ASYMPTOTIC PROPERTIES AND COMPUTATION OF MAXIMUM LIKELIHOOD ESTIMATES IN THE MIXED MODEL OF THE ANALYSIS OF VARIANCE

BY

JOHN JAMES MILLER

TECHNICAL REPORT NO. 12
NOVEMBER 21, 1973

PREPARED UNDER CONTRACT
N00014-67-A-0112-0030 (NR-042-034)
FOR THE OFFICE OF NAVAL RESEARCH

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ASYMPTOTIC PROPERTIES AND COMPUTATION OF MAXIMUM LIKELIHOOD ESTIMATES IN THE MIXED MODEL OF THE ANALYSIS OF VARIANCE

BY

JOHN JAMES MILLER
Stanford University

TECHNICAL REPORT NO. 12

NOVEMBER 21, 1973

PREPARED UNDER THE AUSPICES

OF

OFFICE OF NAVAL RESEARCH CONTRACT # N00014-67-A-0112-0030

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Acknowledgments

I would like to express my sincerest thanks to Professor Theodore W. Anderson for his guidance in the preparation of this dissertation. His insight into this complicated and often frustrating topic was invaluable. I am especially grateful for his patience and for his understanding of difficulties which were beyond anyone's control. I would also like to thank Professors Ingram Olkin and Louis Gordon, who read the manuscript and made helpful suggestions. I thank my friends and family and especially my wife, Ruth Ann, for their patience and encouragement. Thanks also go to Mrs. Anna Corsaro, who typed the rough draft, and to Mrs. Mary Daughaday who so excellently prepared the final manuscript.

TABLE OF CONTENTS

APPENDIX A

CHAPTER 1

INTRODUCTORY REMARKS

1.1. Introduction

The problem considered in this paper is the estimation of the parameters in the mixed model of the analysis of variance by the method of maximum likelihood, assuming normality of the random effects and errors. Both asymptotic properties of such estimators as the size of the design increases and numerical methods for their calculation are considered. The mixed model has been studied for many years. Various methods have been suggested for use in specific cases, but without unified theory applying to all cases. The method of maximum likelihood, which provides such a unified theory, has not been used in the past because the complexity of the likelihood equations in general cases made their solution very difficult. Computers have now made feasible the solution of the likelihood equations; this fact makes the study of the properties of these estimators of interest.

This paper extends certain asymptotic results of H. O. Hartley and J. N. K. Rao (1967) and Whitby (1971) to cover the asymptotic behavior of these estimators in great generality. Both of these previous sets of results have restrictions confining their application to a narrower class of models to which many interesting cases do not belong. In this paper the theory has been extended to cover almost all models. (Certain degenerate cases are not considered.) The

theory presented here applies to both balanced and unbalanced designs[1].

The exact model used is given by Hartley and Rao as

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \underset{\sim}{U}_1\underset{\sim}{b}_1 + \underset{\sim}{U}_2\underset{\sim}{b}_2 + \ldots + \underset{\sim}{U}_{p_1}\underset{\sim}{b}_{p_1} + \underset{\sim}{e} \quad ,$$

where $\underset{\sim}{y}$ is an nx1 vector of observations; $\underset{\sim}{X}$ is an nxp_0 design matrix for the p_0x1 vector of fixed effects $\underset{\sim}{\alpha}$; $\underset{\sim}{U}_i$ is an nxm_i design matrix for the m_ix1 vector of random effects $\underset{\sim}{b}_i$, $i=1,2,\ldots,p_1$; $\underset{\sim}{e}$ is an nx1 random vector of errors. It is assumed that the expected value of each of the random vectors is the zero vector and that the covariance matrix of $\underset{\sim}{b}_i$ is $\sigma_i\underset{\sim}{I}$, $i=1,2,\ldots,p_1$, and the covariance matrix of $\underset{\sim}{e}$ is $\sigma_0\underset{\sim}{I}$, where the identity matrices are of appropriate size. (See the note in Section 1.3 on the use of σ_i instead of σ_i^2 for a variance.) It is further assumed that $\underset{\sim}{b}_1, \underset{\sim}{b}_2, \ldots, \underset{\sim}{b}_{p_1}$, and $\underset{\sim}{e}$ are mutually independent and that each has a multivariate normal distribution. It follows that $\underset{\sim}{y}$ has a multivariate normal distribution with mean $\underset{\sim}{X}\underset{\sim}{\alpha}$ and covariance matrix $\underset{\sim}{\Sigma} \equiv \sigma_0\underset{\sim}{I} + \sigma_1\underset{\sim}{U}_1\underset{\sim}{U}_1' + \sigma_2\underset{\sim}{U}_2\underset{\sim}{U}_2' + \ldots + \sigma_{p_1}\underset{\sim}{U}_{p_1}\underset{\sim}{U}_{p_1}'$. A full description of the model and the basic assumptions about it is given in Section 1.3. The problem is to estimate $\underset{\sim}{\alpha}$ and $\sigma_0, \sigma_1, \ldots, \sigma_{p_1}$ by the method of maximum likelihood and to study the properties of such estimators.

Consistency and asymptotic normality of maximum likelihood estimators are often proved by using a Taylor series expansion of the

1. The author acknowledges his indebtedness to Hartley and Rao for the form of the basic model and to Whitby for the form of Theorems 3.2.1 and 3.3.1.

likelihood equations about the true parameter point (Cramér [1946]). There are often independent, identically distributed observations and, in this case, conditions are placed on the common density function to allow the correct expansions to be made. The asymptotic theory is then carried out by normalizing by some sequence depending on the number of observations. (Usually, but not always, the square root is used; see Whitby [1971:2].) Work has also been done with cases where the observations are not independent, not identically distributed, or neither independent nor identically distributed. (For example, see Silvey [1961].) However, none of the above theory can be directly applied to this problem. One generally does not think of an analysis of variance as a sequence of observations but as an experiment; that is, one thinks of an analysis of variance as one observation of a vector of variables constituting an experiment. Asymptotic theory is then usually carried out on a "conceptual sequence of experiments" for which the size of the entire design becomes infinite in some orderly way. The work done in this area of the analysis of variance does not apply in all cases either. The problem is that often the estimate of each parameter requires a different normalizing sequence; this problem is discussed further below.

The results presented here are extensions of the work of Whitby in the following sense. Whitby proved theorems dealing with a general maximum likelihood estimation problem. He considered estimation of several parameters but his proofs depend on the normalizing sequence (he calls it c_n) being the same for each parameter and on the norm of

a pxl vector $\underset{\sim}{x}$ being $\|\underset{\sim}{x}\| = \left(\Sigma_{i=1}^{p} x_i^2\right)^{\frac{1}{2}}$. He mentioned that different normalizing sequences may be necessary but gave no indication of how they are obtained or why they might be necessary. The theorems presented here are much more general. They allow any legitimate vector norm to be used and they allow the estimate of each parameter to have its own normalizing sequence. To see that such an extension may be necessary in an analysis of variance, one need only consider the balanced two-way model. The sufficient statistics in this model are the grand mean and several sums of squares. The sums of squares are a set of independent chi-square random variables with degrees of freedom which will increase at different rates (See Sections 6.1 and 6.2.); any analysis of the maximum likelihood estimators (which of course are functions of the sufficient statistics) must take account of these differences.

The freedom to allow the estimate of each parameter to have its own normalizing sequence is achieved by the artifice of building the normalizing sequence into the parameter, obtaining a set of sequences of parameters. The basic asymptotic theorem, Theorem 3.3.1, deals only with such sequences of parameters. It is quite general; in fact, most of the usual asymptotic results about maximum likelihood estimation can be derived as special cases of Theorem 3.3.1. In Chapter 4 Theorem 3.3.1 is used to prove the consistency and asymptotic normality of the maximum likelihood estimators in the mixed model of the analysis of variance. This is done by translating back from the properties of estimators of a set of sequences of parameters to a

sequence of estimates of a set of parameters.

One advantage of the use of the above method of proof for Theorem 3.3.1 and its application in Theorem 4.4.1 is that the problems incurred when the wrong normalizing sequence is used are clearly located. These problems are easily illustrated by considering the simple case of correctly normalizing $\overline{X}_n$, the arithmetic mean of n independent, identically distributed random variables, each with mean zero and finite nonzero variance. If the normalized estimate is $Y_n = n^{\frac{1}{2}+\epsilon}\ \overline{X}_n$, then only for $\epsilon=0$ will a limiting normal distribution be obtained. If $\epsilon < 0$ the normalizing sequence is "too small" and the Y_n converge to a degenerate (point) distribution at zero. If $\epsilon > 0$ the normalizing sequence is "too large" and the distributions of Y_n "blow up" to a distribution having atoms at plus and minus infinity. This phenomenon has sometimes been described in the following manner: if $\epsilon < 0$ the asymptotic variance is zero and if $\epsilon > 0$ the asymptotic variance is infinite. While the first descriptions of the phenomenon are technically more accurate, the second descriptions do point out a way of locating the problem. In the analysis of variance model the problems manifest themselves in the matrix $\underset{\sim}{J}$ defined in Section 4.3.

The matrix $\underset{\sim}{J}$ is the limit of the matrix of expected values of second derivatives of the log-likelihood. This matrix has had the normalizing sequences built into it. Only if the normalizing sequence for each parameter is of precisely the right order of magnitude will $\underset{\sim}{J}$ be positive definite. If the sequence for some parameter is "too

small" the limit of the appropriate diagonal element will be infinite. If the sequence is "too large" the entire row and column of $\underset{\sim}{J}$ associated with that parameter will be zero. Since $\underset{\sim}{J}^{-1}$ is the asymptotic covariance matrix it is easily seen that results analagous to the case for $\overline{X}$ occur in the model under study. The problems of zero or infinite "asymptotic variances" manifest themselves in the nonexistence of the limits forming $\underset{\sim}{J}$ or the fact that $\underset{\sim}{J}$ is not positive definite. However, the proofs of Theorems 3.3.1 and 4.4.1 point out further what goes wrong. The proof of Theorem 3.3.1 breaks down completely if $\underset{\sim}{J}$ is not positive definite. Proofs of the lemmae used to prove Theorem 4.4.1 also break down if the normalizing sequences are not of the correct orders of magnitude. The assumptions of Section 4.2 insure that for the sequence of designs considered in Theorem 4.4.1 the matrix $\underset{\sim}{J}$ will be positive definite.

The results presented here are also extension of the work of H. O. Hartley and J. N. K. Rao. Hartley and Rao (1967:101) make the following assumption about the asymptotic behavior of the design matrices $\underset{\sim}{U}_i$: Every column of each $\underset{\sim}{U}_i$ may contain at most a finite number of nonzero elements. The two-way balanced model mentioned above illustrates that this assumption rules out any sort of balanced-crossed layout. (See Section 6.1.) The effect of this assumption is, in fact, to assure that the estimates of all the σ_i can be normalized by the same normalizing sequence. As noted above, the results presented here are not so restrictive. In fact, any sequence of designs that

might be considered as an actual sequence of experiments is covered by these results; the assumptions used here may appear restrictive but they merely rule out cases that are useful only as counterexamples to theorems and not as actual design sequences.

Hartley and Rao went on in their paper to assert that under their assumptions, the maximum likelihood estimators are consistent and asymptotically efficient, although they did not make clear what they meant by the latter term. They sketched but did not give details of a method of proof. In fact the details constitute the difficult part of the proof. With a great deal of effort, and after corrections are made to their assumptions, their method of proof (with details) yields the claimed results. (If the assumptions are taken as written, counterexamples to the theorems can be found.) A more detailed discussion of their proof is given in Appendix D. The methods used here require no less effort but cover all interesting cases.

Thus far only asymptotic theory concerning the maximum likelihood estimates has been discussed. The numerical computation of the estimates is also a fruitful area for study. One problem in this case is the problem of negative estimates. The maximum likelihood estimate of a variance can never be a negative number; such a point would not be in the parameter space and would be ineligible for a maximum likelihood estimate. Thus some sort of truncation procedure must be used to insure nonnegative estimates. The procedure used when the estimates are obtained as solutions to the likelihood equations is given in

Section 1.3. The problem of negative estimates is also discussed in Chapter 2 and in Section 5.6. It should be pointed out that truncation does not affect the asymptotic theory because the true parameter is assumed to be an interior point of the parameter space so that all the true variances are positive. Since the estimates are consistent, truncation ceases to be a problem. Asymptotic theory when the true parameter point is on the boundary has not been developed for this model; in some simple boundary cases, asymptotic distributions occur which are definitely not normal. Thus some further extension of methods of proof will be required in these cases.

The actual numerical procedures for the computation of the estimates are discussed in Chapter 5. An iterative procedure suggested by Anderson (1971b), (1973) is compared with a procedure suggested by H. O. Hartley and J. N. K. Rao (1967) and implemented by Vaughn (1970). The procedure of Anderson was found to be computationally much more efficient than the Hartley, Rao, Vaughn procedure in a Monte Carlo study. J. N. K. Rao (1973) has pointed out that the Anderson procedure is in effect the method of scoring. A computer program developed to implement the Anderson procedure is discussed in Appendix C.

1.2. Notation and Conventions Used

The following notation will be used throughout this paper. All vectors are column vectors and are underscored with the symbol "$\sim$" to represent boldface type. Matrices are also underscored in the same manner. With three exceptions ($\underset{\sim}{G}$, $\underset{\sim}{R}$, and sometime $\underset{\sim}{Y}$) vectors are represented by small Latin or Greek letters; matrices are always represented by capital Latin or Greek letters. All vectors belong to the space R^m (the Cartesian product of R, the real line, with itself m times) of the appropriate dimension. A vector norm is denoted $\|\cdot\|$ and the matrix norm it induces as a natural norm is also denoted by $\|\cdot\|$; i.e., $\|\underset{\sim}{A}\| = \sup_{\underset{\sim}{x} \neq \underset{\sim}{0}} \frac{\|\underset{\sim}{A}\underset{\sim}{x}\|}{\|\underset{\sim}{x}\|}$. If $\underset{\sim}{A}$ is an nxn matrix, $tr(\underset{\sim}{A}) = \Sigma_{i=1}^{n} a_{ii}$ is the trace of $\underset{\sim}{A}$, $\lambda_j(\underset{\sim}{A})$ is the j^{th} characteristic root of $\underset{\sim}{A}$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, $|\underset{\sim}{A}|$ is the determinant of $\underset{\sim}{A}$, and if $\underset{\sim}{A}$ is non-singular $\underset{\sim}{A}^{-1}$ is the inverse of $\underset{\sim}{A}$ and $\underset{\sim}{A}^{-t} = (\underset{\sim}{A}')^{-1} = (\underset{\sim}{A}^{-1})'$.

If $\underset{\sim}{f}(\underset{\sim}{x})$ is a px1 vector function of the px1 vector $\underset{\sim}{x}$, then the Jacobian of $\underset{\sim}{f}$, $\underset{\sim}{J}_{\underset{\sim}{f}}(\underset{\sim}{x})$ is a pxp matrix function of $\underset{\sim}{x}$ defined by $[\underset{\sim}{J}_{\underset{\sim}{f}}(\underset{\sim}{x})]_{ij} \equiv \frac{\partial[\underset{\sim}{f}(\underset{\sim}{x})]_i}{\partial x_j}$. It then follows that if $\underset{\sim}{f}(\underset{\sim}{x}) = \underset{\sim}{A}^{-1}\underset{\sim}{g}(\underset{\sim}{x})$, where $\underset{\sim}{f}$ and $\underset{\sim}{g}$ are px1 vector functions of the px1 vector $\underset{\sim}{x}$ and $\underset{\sim}{A}$ is pxp nonsingular, $\underset{\sim}{J}_{\underset{\sim}{f}}(\underset{\sim}{x}) = \underset{\sim}{A}^{-1}\underset{\sim}{J}_{\underset{\sim}{g}}(\underset{\sim}{x})$. A frequently used notation will be $\underset{\sim}{G}(\underset{\sim}{\psi},\underset{\sim}{Y})$, where $\underset{\sim}{G}$ is a px1 vector function of a px1 variable $\underset{\sim}{\psi}$ and a random variable $\underset{\sim}{Y}$ ($\underset{\sim}{Y}$ may also be a vector). Analyses will then be performed on $\underset{\sim}{G}$ relative to $\underset{\sim}{\psi}$ for fixed $\underset{\sim}{Y}$ and $\underset{\sim}{Y}$ may be required to belong to some set in its probability space (which set will have large

probability). For such a $\underset{\sim}{G}$ the p×p Jacobian matrix $\underset{\sim}{J}_{\underset{\sim}{G}}(\underset{\sim}{\psi},\underset{\sim}{Y})$ is defined $[\underset{\sim}{J}_{\underset{\sim}{G}}(\underset{\sim}{\psi},\underset{\sim}{Y})]_{ij} \equiv \dfrac{\partial[\underset{\sim}{G}(\underset{\sim}{\psi},\underset{\sim}{Y})]_i}{\partial\psi_j}$. Another notation which will be used may seem confusing at first. $\underset{\sim}{\psi}_n$ is often used as a statistical parameter and $\underset{\sim}{\psi}$ as an argument in a function in the same expression; the notation $\left.\dfrac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n}$ may appear, where λ is a scalar function of $\underset{\sim}{y}$ and $\underset{\sim}{\psi}$ and $\underset{\sim}{y}$ is a random variable. This expression contains $\underset{\sim}{\psi}$, the variable which is differentiated with respect to in the function λ, and also $\underset{\sim}{\psi}_n$, a parameter, one of a sequence of parameters under consideration. This dual role of $\underset{\sim}{\psi}$ parallels the use of dummy variables in an integration (e.g. $f(x) = \int_{x_0}^{x} g(x)dx$). Differentiation by a subvector will be as follows: Let $\underset{\sim}{\psi}' = (\underset{\sim}{\beta}', \underset{\sim}{\tau}')$ then $\left.\dfrac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\beta_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n}$ or $\left.\dfrac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\tau_i}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n}$ may be written if that is more convenient than using a complicated subscript scheme. Vector derivatives may also be used where appropriate; for example, $\dfrac{\partial\lambda}{\partial\underset{\sim}{\beta}}$ where $\underset{\sim}{\beta}$ is p_0×1 yields a p_0×1 vector of derivatives and $\dfrac{\partial^2\lambda}{\partial\underset{\sim}{\beta}\partial\underset{\sim}{\beta}'}$ and yields a p_0×p_0 matrix of second derivatives.

Linear spaces are denoted by script letters and are column spaces; that is, $\mathcal{L}(\underset{\sim}{X})$ is the linear space formed by all linear combinations of the columns of $\underset{\sim}{X}$. The symbol $\oplus$ denotes direct sum; if $\mathcal{X} = \mathcal{L}(\underset{\sim}{X})$ and

$\mathcal{Y} = \mathcal{L}(\underset{\sim}{Y})$ then $\mathcal{Z} = \mathcal{X} \oplus \mathcal{Y} \equiv \{\underset{\sim}{z} | \underset{\sim}{z} = \underset{\sim}{x} + \underset{\sim}{y},\ \underset{\sim}{x} \in \mathcal{X},\ \underset{\sim}{y} \in \mathcal{Y}\}$. Spheres of radius r about $\underset{\sim}{x}_0$ are represented by $S_r(\underset{\sim}{x}_0) \equiv \{\underset{\sim}{y} |\ \|\underset{\sim}{y} - \underset{\sim}{x}_0\| < r\}$ for whatever norm $\|\cdot\|$ is being used. If $S \subset R^p$, then $\overline{S}$ is the closure of S.

Convergence in distribution is represented by $\xrightarrow{d}$ and convergence in probability by $\xrightarrow{p}$. Both of these concepts may be used with vectors or matrices. For instance, "$\underset{\sim}{A}_n(\underset{\sim}{Y}_n) \xrightarrow{p} \underset{\sim}{A}$", where $\underset{\sim}{A}_n(\underset{\sim}{Y}_n)$ is a pxp matrix function of a random variable $\underset{\sim}{Y}_n$ and $\underset{\sim}{A}$ is a pxp constant matrix, means $P\{\|\underset{\sim}{A}_n(\underset{\sim}{Y}_n) - \underset{\sim}{A}\| > \delta\} \to 0$ as $n \to \infty$. Since both vector and matrix norms are continuous functions of their elements, it is sufficient to prove convergence for each element of the vector or matrix separately.

A p dimensional multivariate normal random vector is denoted $\underset{\sim}{\mathcal{N}}_p(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$, where $\underset{\sim}{\mu}$ is its expected value and $\underset{\sim}{\Sigma}$ is its covariance matrix. For a univariate normal, the subscript is dropped. χ^2_p is a chi-square random variable with p degrees of freedom. The symbol "$\sim$" means "distributed as"; thus $X \sim \chi^2_p$ means the random variable X has a chi-square distribution with p degrees of freedom.

One practice followed in this paper which the reader might find confusing is the notation of dependence on n. "$n \to \infty$" is used to denote that the size of the entire design becomes infinite ($\underset{\sim}{y}$ is nx1). All other elements of the problem, including the sizes of vectors and matrices, depend on n; only p_0 and p_1, the number of parameters in the model, remain fixed. The dependence on n does not always appear in the notation. It is suppressed when it is obvious that such dependence exists. When it is not suppressed, it is usually to emphasize the dependence on n. The reader, being forewarned, should not be disturbed

by the seeming inconsistency; as one becomes familiar with the topic, the inconsistency disappears.

This paper is divided into seven chapters labeled 1-7 and four appendices labeled A-D. When a chapter is divided into sections, these sections are labeled 1,2,... and are prefixed by the chapter label (e.g. Section 1.2, Section A.3). Section A.4 is further divided into five subsections labeled A.4.1-A.4.5. Theorems, lemmae, propositions, and assumptions are numbered consecutively within sections and prefixed by the section label (e.g. Assumption 1.3.5, Proposition A.3.4). If there is only one theorem in a section it is still labeled as the first theorem in the section (e.g. Theorem 3.2.1). In the case of the subsections A.4.1-A.4.5, the lemmae appearing in these subsections are labeled A.4.1-A.4.5 instead of A.4.1.1-A.4.5.1 because the first set of numbers relates to Section A.4 as a whole and that is the rule used to number these lemmae.

1.3. Basic Analysis of Variance Model and Assumptions About It

The basic model used will be the mixed model analysis of variance, which can be written as

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \underset{\sim}{U}_1\underset{\sim}{b}_1 + \underset{\sim}{U}_2\underset{\sim}{b}_2 + \dots + \underset{\sim}{U}_{p_1}\underset{\sim}{b}_{p_1} + \underset{\sim}{e}$$

where

$\underset{\sim}{y}$ is an $n \times 1$ vector of observations;

$\underset{\sim}{X}$ is an $n \times p_0$ matrix of known constants;

$\underset{\sim}{\alpha}$ is a $p_0 \times 1$ vector of unknown constants;

$\underset{\sim}{U}_i$ is an $n \times m_i$ matrix of known constants, $i=1,2,\dots,p_1$;

$\underset{\sim}{b}_i$ is an $m_i \times 1$ random vector, $i=1,2,\dots,p_1$;

$\underset{\sim}{e}$ is an $n \times 1$ random vector.

Thus $\underset{\sim}{X}$ is the design matrix for the fixed effects and the $\underset{\sim}{U}_i$ are the design matrices for the random effects $\underset{\sim}{b}_i$. Let $\underset{\sim}{G}_i = \underset{\sim}{U}_i\underset{\sim}{U}_i'$, $i=1,2,\dots,p_1$, and $\underset{\sim}{G}_0 = \underset{\sim}{I}_n$. The following assumptions are made about the model:

ASSUMPTION 1.3.1. <u>The random vectors $\underset{\sim}{b}_1, \underset{\sim}{b}_2, \dots, \underset{\sim}{b}_{p_1}$, $\underset{\sim}{e}$ are mutually independent, with $\underset{\sim}{e} \sim \mathcal{N}_n(\underset{\sim}{0}, \sigma_0 \underset{\sim}{I}_n)$ and $\underset{\sim}{b}_i \sim \mathcal{N}_{m_i}(\underset{\sim}{0}, \sigma_i \underset{\sim}{I}_{m_i})$,</u> $i=1,2,\dots,p_1$.[2]

ASSUMPTION 1.3.2. <u>The matrix $\underset{\sim}{X}$ has full rank p_0.</u>

ASSUMPTION 1.3.3. $n \geq p_0 + p_1 + 1$.

ASSUMPTION 1.3.4. <u>The partitioned matrix $[\underset{\sim}{X} \vdots \underset{\sim}{U}_i]$ has rank greater than p_0,</u> $i=1,2,\dots,p_1$.

2. This differs from the usual convention of using σ_i^2. σ_i is used as a <u>variance</u> to avoid writing many squares. This also follows the notation of Anderson (1969), (1970), (1971b), (1973).

ASSUMPTION 1.3.5. <u>The matrices $\underset{\sim}{G}_0, \underset{\sim}{G}_1, \ldots, \underset{\sim}{G}_{p_1}$ are linearly independent; that is, $\Sigma_{i=0}^{p_1} \sigma_i \underset{\sim}{G}_i = \underset{\sim}{0}$ implies</u> $\sigma_i = 0$, $i=0,1,\ldots,p_1$.

These assumptions are sufficient to find the estimates; however, one further assumption about the $\underset{\sim}{U}_i$ will be made.

ASSUMPTION 1.3.6. <u>The matrix $\underset{\sim}{U}_i$ consists only of zeros and ones and there is exactly one 1 in each row and at least one 1 in each column</u>, $i=1,2,\ldots,p_1$.

The above assumptions can be explained in analysis of variance terms. Assumption 1.3.1 is the usual assumption of the independence and normality of the random effects. Assumption 1.3.2 can always be satisfied by a suitable reparameterization of the problem. Assumption 1.3.3 says there are at least as many observations as parameters. Assumption 1.3.4 says that the fixed effects are not confounded with any of the random effects. Assumption 1.3.5 says that the random effects are not confounded with each other. Assumption 1.3.6 just says that the $\underset{\sim}{U}_i$ are standard design matrices and it has three consequences. $\underset{\sim}{U}_i' \underset{\sim}{U}_i = \underset{\sim}{D}_i$, an $m_i \times m_i$ nonsingular diagonal matrix; $\underset{\sim}{U}_i$ has full rank m_i; and $m_i \leq n$.

It follows from the above assumptions that $\underset{\sim}{y}$ has a normal distribution with

$$\mathscr{E}(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\alpha} ,$$

$$\begin{aligned} \text{Cov}(\underset{\sim}{y}) &= \mathscr{E}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha})(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha})' \\ &= \sigma_0 \underset{\sim}{I}_n + \sigma_1 \underset{\sim}{U}_1 \underset{\sim}{U}_1' + \ldots + \sigma_{p_1} \underset{\sim}{U}_{p_1} \underset{\sim}{U}_{p_1}' \\ &= \sum_{i=0}^{p_1} \sigma_i \underset{\sim}{G}_i \equiv \underset{\sim}{\Sigma} . \end{aligned}$$

Thus $\underset{\sim}{y} \sim \mathcal{N}_n(\underset{\sim}{X}\underset{\sim}{\alpha},\underset{\sim}{\Sigma})$. The problem considered here is to observe $\underset{\sim}{y}$ and estimate $\underset{\sim}{\alpha}$, $\sigma_0,\sigma_1,\ldots,\sigma_{p_1}$ by the method of maximum likelihood.

The parameter space is defined as follows: Let $p=p_0+p_1+1$. $\Theta \subset R^p$ is the parameter space; if $\underset{\sim}{\theta} \in \Theta$ then $\underset{\sim}{\theta}' = (\underset{\sim}{\alpha}', \underset{\sim}{\sigma}')$, where $\underset{\sim}{\alpha} = (\alpha_1,\alpha_2,\ldots,\alpha_{p_0})'$ and $\underset{\sim}{\sigma} = (\sigma_0,\sigma_1,\ldots,\sigma_{p_1})'$. The restrictions which form Θ are $\Theta = \{\underset{\sim}{\theta} \in R^p | \underset{\sim}{\theta}' = (\underset{\sim}{\alpha}',\underset{\sim}{\sigma}'); \underset{\sim}{\alpha} \in R^{p_0}; \sigma_0 > 0; \sigma_i \geq 0, i=1,2,\ldots,p_1\}$. If the likelihood function of $\underset{\sim}{y}$ and $\underset{\sim}{\theta}$ is $L(\underset{\sim}{y},\underset{\sim}{\theta})$ it follows from the multivariate normal density that

$$\lambda(\underset{\sim}{y},\underset{\sim}{\theta}) \equiv \log L(\underset{\sim}{y},\underset{\sim}{\theta})$$

$$= -\tfrac{1}{2}n(\log 2\pi) - \tfrac{1}{2}\log|\underset{\sim}{\Sigma}| - \tfrac{1}{2}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha})'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}).$$

Differentiation[3] of λ by $\underset{\sim}{\alpha}$ and $\underset{\sim}{\sigma}$ leads to the following equations which must be solved simultaneously for $\hat{\underset{\sim}{\alpha}}$ and $\hat{\underset{\sim}{\sigma}}$:

$$\left[\underset{\sim}{X}'\left(\sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j\right)^{-1}\underset{\sim}{X}\right]\underset{\sim}{\alpha} = \underset{\sim}{X}'\left(\sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j\right)^{-1}\underset{\sim}{y} \quad ,$$

$$\mathrm{tr}\left(\sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j\right)^{-1}\underset{\sim}{G}_i = \mathrm{tr}\left(\sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j\right)^{-1}\underset{\sim}{G}_i\left(\sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j\right)^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha})(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha})',$$

$$i=0,1,\ldots,p_1.$$

These equations can be abbreviated

$$[\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{X}]\underset{\sim}{\alpha} = \underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{y} \, ,$$

$$\mathrm{tr}\,\underset{\sim}{\Sigma}^{-1}\underset{\sim}{G}_i = \mathrm{tr}\,\underset{\sim}{\Sigma}^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha})(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha})';\, i=0,1,\ldots,p_1 \, ,$$

where $\underset{\sim}{\Sigma}$ is a function of $\underset{\sim}{\sigma}$.

3. Such maximization by differentiating is justified in Anderson (1970).

As was noted in Section 1.1, the maximum likelihood estimates of the variances cannot be negative, but the solutions of the likelihood equations may be negative. Thus the following truncation scheme is necessary. If the solution of the likelihood equations yields a negative estimate of σ_i for $i \in S$ (where S is some set of indices) then solve the following set of equations:

$$\frac{\partial\lambda}{\partial\underset{\sim}{\alpha}} = \underset{\sim}{0} \quad ,$$

$$\frac{\partial\lambda}{\partial\sigma_i} = 0 \quad , \quad i \notin S \ ,$$

$$\sigma_i = 0 \quad , \quad i \in S \ .$$

If a negative estimate occurs for some other σ_i, add its index to the set S and repeat the above procedure. Continue in this manner until no negative estimates are obtained. As is pointed out in Section 5.6 this is easily done for this model.

The computational methods used to solve the likelihood equations are discussed in Chapter 5. The asymptotic properties of such solutions are discussed in Chapter 4.

CHAPTER 2

REVIEW OF PAST LITERATURE

The subject of variance component estimation has been considered for some time. An excellent overview of the present state of the art may be found in Searle (1971). He described the development of many different techniques used and gave copious references. A history of the development of the entire field will not be given here; instead a short review of the literature immediately pertinent to this paper is presented.

Theoretical properties of methods of estimation and testing of variance components have been considered by Herbach (1959), who considered the one and two way balanced layouts and proved optimality results for the usual analysis of variance tests and also derived the likelihood ratio tests in these cases. Graybill and Hultquist (1961), Hultquist and Graybill (1965), and Hultquist and Atzinger (1973) considered the balanced models with respect to minimal sufficient statistics and proved various optimality results as well as deriving certain likelihood equations; some of the results of Hultquist and Atzinger overlap some of those of Anderson (1970) described below.

Several authors have considered the problem of a model where the covariance matrix has a special structure. Wilks (1946) considered the intraclass correlation coefficient model. Olkin and Press (1969) studied the circular stationary model. Srivastava (1966) and Srivastava

and Maik (1967) considered a more general model where the covariance matrix has linear structure and the matrices $\underset{\sim}{G}_0, \underset{\sim}{G}_1, \ldots, \underset{\sim}{G}_{p_1}$ have special properties. All the above authors derived likelihood ratio tests (using several diverse techniques) for the particular model under study. Anderson (1969), (1970), (1971b), (1973) also studied models where the covariance matrix has linear structure. Anderson's method of analysis enabled all the above cases to be considered within one unified framework.

The model Anderson used is $\underset{\sim}{y} \sim \underset{\sim}{\mathcal{N}}_p(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$, where $\underset{\sim}{\mu} = \underset{\sim}{X}\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma} = \sum_{i=0}^{m} \sigma_i \underset{\sim}{G}_i$. The model assumed here is a special case of this. Anderson derived the likelihood equations and showed how they can be simplified in certain cases. (See Section 5.2.) He gave conditions for estimability of the parameters and suggested several methods for the solution of the likelihood equations. The method studied in this paper was advanced in (1971b). In (1969), he derived and gave properties of the likelihood ratio tests of hypotheses about $\underset{\sim}{\Sigma}$ and the σ_i. He also proved that the maximum likelihood estimators in this case are consistent and asymptotically efficient as the entire process is replicated (that is, as repeated observations are taken on $\underset{\sim}{y}$) and he derived the asymptotic covariance matrix.

H. O. Hartley and J. N. K. Rao (1967) analyzed maximum likelihood estimation in the mixed model of the analysis of variance, the model used in this paper. They gave five rationale for using the method of maximum likelihood in this case, which are paraphrased here.

a) Computers make easy solution of the likelihood equations possible.

b) This technique can be applied to any model, balanced or unbalanced.

c) The technique has large sample optimality properties.

d) Maximum likelihood estimates are always functions of the minimal sufficient statistics.

e) The maximum likelihood estimates of the variance components are always positive.

Several comments can be made about these rationale. A technique (similar to Anderson's) is proposed in this paper which is different than Hartley and Rao's for solving the likelihood equations. This technique does not guarantee nonnegative estimates, but can easily be modified so that only nonnegative estimates are finally arrived at.(See Chapter 5.) (Many writers have considered the problem of negative estimates; see Searle [1971:22] for a good summary.) Hartley and Vaughn (1972) developed a computer program to implement the algorithm of Hartley and Rao. (The computer program to implement the algorithm proposed here is discussed in Appendix C.) The large sample optimality referred to above was proved by Hartley and Rao only for a limited set of designs in which the number of observations at any particular level of any random factor must remain bounded. Such an assumption rules out even so simple a model as the two-way crossed layout random effects model. In this paper the optimality results are extended to cover almost all interesting cases.

The discussion of asymptotic results has, as a rule, been confined to independent, identically distributed observations. The basic techniques were expounded by Cramér (1946), Wald (1949) and Wolfowitz (1949). As noted in the introduction, these techniques do not apply to the model under consideration here. Silvey (1961) discussed asymptotic results for sequences of dependent observations but again his work cannot be applied in this case. Whitby (1971) considered estimation for the generalized beta distribution; although he was also considering sequences of independent, identically distributed random variables, he did prove some general theorems which have been extended to cover the analysis of variance model of this paper. Whitby used only one normalizing sequence for all the estimates. The theorems presented here allow a different normalizing sequence for the estimate of each parameter. This is done by the artifice of building the normalizing sequence right into the parameter. The result is a sequence of parameters, the estimates of which have certain properties; these properties can then easily be translated to properties of the desired estimators. (See Chapter 4.)

This is only a cursory review of the subject of maximum likelihood estimation in the analysis of variance. For a more detailed study, the reader is referred to the papers of Searle (1971); Anderson (1971b), (1973); Hartley and Rao (1967); and Whitby (1971).

CHAPTER 3

BASIC THEOREMS ON ASYMPTOTIC BEHAVIOR

3.1. Introduction

In this chapter two theorems are presented which will be used to prove the asymptotic properties of maximum likelihood estimators in the analysis of variance. Theorem 3.2.1 is a form of inverse function theorem and is used to prove Theorem 3.3.1. Theorem 3.3.1 is a very general theorem concerning asymptotic theory and is used to prove Theorem 4.4.1, the main asymptotic result of this paper. However, where Theorem 3.2.1 has little intrinsic interest except for the mathematics of its proof, Theorem 3.3.1 has applications beyond that of a lemma for Theorem 4.4.1. Theorem 3.3.1 has wide applicability and can be used to prove most of the standard asymptotic results concerning roots of the likelihood equation.

Theorem 3.2.1 is a form of inverse function theorem. It concerns roots of the vector equation $\underset{\sim}{G}(\underset{\sim}{x}) = \underset{\sim}{0}$. When $\underset{\sim}{G}$ has the form $\underset{\sim}{G}(\underset{\sim}{x}) = \underset{\sim}{a} + \underset{\sim}{A}(\underset{\sim}{x}-\underset{\sim}{x}_0) + \underset{\sim}{r}(\underset{\sim}{x})$, $\underset{\sim}{a}$ and $\underset{\sim}{r}$ are small relative to $\underset{\sim}{A}$, and certain continuity and differentiability conditions are met, then there is a root of $\underset{\sim}{G}(\underset{\sim}{x}) = \underset{\sim}{0}$ near $\underset{\sim}{x}_0$. Similar theorems are often used in proving results about the roots of the likelihood equations. The method of proof of Theorem 3.2.1 is patterned on the proof of The Inverse Function Theorem (Theorem 9.17) of Rudin (1964:193-195). The form of the theorem is patterned on a lemma (Lemma 3.1) of Whitby (1971:8-9). Whitby based his proof on the proof in the first edition

of Rudin's book; Rudin's first proof (and hence Whitby's) required that the norm of a vector be the Euclidean norm, $\|\underset{\sim}{x}\| = \left(\Sigma_{i=1}^{p} x_i^2\right)^{\frac{1}{2}}$. The proof given here is slightly more general in that it allows any vector norm to be used. (See Issacson and Keller [1966:3-4] for the definition of a vector norm.) The assumptions of Theorem 3.2.1 are stated somewhat differently than those of Whitby's Lemma 3.1 in order to facilitate the different method of proof used here.

Theorem 3.3.1 contains powerful asymptotic results and has intrinsic interest because of its wide applicability. Theorem 3.3.1 concerns a sequence of estimators of a sequence of constants (or constant vectors). The sequence of estimators becomes close (in a well defined sense) to the sequence of constants with high probability; furthermore, the sequence of estimates is asymptotically normal. As seen in Chapter 4, this sequence of estimates can easily be translated into a sequence of consistent and asymptotically normal estimates of the parameters in the analysis of variance problem. The basic setup and method of proof of Theorem 3.3.1 are as follows: For each n (of a sequence of values of n increasing to infinity) there is a vector function $\underset{\sim}{G}_n(\underset{\sim}{\psi}_n, \underset{\sim}{Y}_n)$ of a vector of parameters $\underset{\sim}{\psi}_n$ and a random vector $\underset{\sim}{Y}_n$. ($\underset{\sim}{G}_n$ will be the likelihood equations.) Then for each $\underset{\sim}{Y}_n$ in a certain set of $\underset{\sim}{Y}_n$ values, it is shown that $\underset{\sim}{G}_n(\underset{\sim}{\psi}_n, \underset{\sim}{Y}_n)$, as a function of $\underset{\sim}{\psi}_n$, satisfies the conditions of Theorem 3.2.1; thus there is a root $\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n)$ of $\underset{\sim}{G}_n(\underset{\sim}{\psi}_n, \underset{\sim}{Y}_n) = \underset{\sim}{0}$ near $\underset{\sim}{\psi}_{0n}$. ($\underset{\sim}{\psi}_{0n}$ will be analogous to the "true" parameter point.) If n is large enough, the set of $\underset{\sim}{Y}_n$ values will have

large probability. It is also shown that $\hat{\psi}_n(Y_n) - \psi_{0n}$ converges in distribution to a multivariate normal distribution.

The wide applicability of Theorem 3.3.1 results from the fact that a sequence of estimates $\{\hat{\psi}_n(Y_n)\}$ of a sequence of parameters $\{\psi_{0n}\}$ is considered instead of a sequence of estimates $\{\hat{\theta}_n(Y_n)\}$ of a single parameter θ_0. The estimates of the single parameter θ_0 will require normalizing sequences and any proofs of asymptotic properties must take explicit account of these normalizing sequences. In Theorem 3.3.1 the normalizing factors are built into the estimates and parameters and are not explicitly mentioned. (In Chapter 4 it is shown how ψ_n is obtained from θ by multiplying each element of θ by the appropriate normalizing factor.) The fact that the normalizing sequences are not specifically mentioned in Theorem 3.3.1 allows it to be used in the analysis of variance (where the estimate of each parameter may require a different normalizing sequence), in a case where the estimate of each parameter can be properly normalized by the square root of the number of observations, or in almost any other case of maximum likelihood estimation. For instance, Theorem 3.3.1 can easily be applied to yield consistency and asymptotic normality in the case of independent, identically distributed observations given in most textbooks (e.g. Cramér [1946]).

The precise statements and proofs of Theorems 3.2.1 and 3.3.1 will be given in the following two sections.

3.2. An Inverse Function Theorem

The theorem in this section is a form of inverse function theorem. The essence of the theorem is that when a function $\underset{\sim}{G}(\underset{\sim}{x})$ is of the form $\underset{\sim}{G}(\underset{\sim}{x}) = \underset{\sim}{a} + \underset{\sim}{A}(\underset{\sim}{x}-\underset{\sim}{x}_0) + \underset{\sim}{r}(\underset{\sim}{x})$ and $\underset{\sim}{a}$ and $\underset{\sim}{r}$ are small relative to $\underset{\sim}{A}$, then there is a root of $\underset{\sim}{G}(\underset{\sim}{x}) = \underset{\sim}{0}$ near $\underset{\sim}{x}_0$. The norms, spheres, and Jacobians used below are defined in Section 1.2. The proof of this theorem is patterned on the proof of Theorem 9.17 of Rudin (1964:193-195).

THEOREM 3.2.1. <u>Let</u> $\underset{\sim}{G}(\underset{\sim}{x})$ <u>be a pX1 vector valued function of a pX1 vector</u> $\underset{\sim}{x}$. <u>Let</u> $\underset{\sim}{a}$ <u>be a pX1 constant vector,</u> $\underset{\sim}{A}$ <u>a pXp constant matrix and</u> $\underset{\sim}{r}(\underset{\sim}{x})$ <u>a pX1 vector valued function of</u> $\underset{\sim}{x}$. <u>Suppose there exists</u> $\eta > 0$ <u>such that</u>

i) $\underset{\sim}{G}(\underset{\sim}{x}) = \underset{\sim}{a} + \underset{\sim}{A}(\underset{\sim}{x}-\underset{\sim}{x}_0) + \underset{\sim}{r}(\underset{\sim}{x})$ <u>for all</u> $\underset{\sim}{x} \in S_{4\eta}(\underset{\sim}{x}_0)$,

ii) $|\underset{\sim}{A}| \neq 0$,

iii) $\|\underset{\sim}{A}^{-1}\,\underset{\sim}{a}\| < \eta$,

iv) $\|\underset{\sim}{A}^{-1}\,\underset{\sim}{r}(\underset{\sim}{x})\| < \frac{1}{4}\eta$ <u>for all</u> $\underset{\sim}{x} \in S_{4\eta}(\underset{\sim}{x}_0)$,

v) $\underset{\sim}{G}(\underset{\sim}{x})$ <u>is continuously differentiable in</u> $S_{4\eta}(\underset{\sim}{x}_0)$,

vi) <u>For</u> $\underset{\sim}{f}(\underset{\sim}{x}) = \underset{\sim}{A}^{-1}\underset{\sim}{G}(\underset{\sim}{x})$

$$= \underset{\sim}{A}^{-1}\,\underset{\sim}{a} + (\underset{\sim}{x}-\underset{\sim}{x}_0) + \underset{\sim}{A}^{-1}\,\underset{\sim}{r}(\underset{\sim}{x})$$

<u>and</u> $\underset{\sim}{x}_1 = \underset{\sim}{x}_0 - \underset{\sim}{A}^{-1}\,\underset{\sim}{a}$,

a) $\underset{\sim}{J}_{\underset{\sim}{f}}(\underset{\sim}{x}_1) \equiv \underset{\sim}{B}$ is nonsingular,

b) $\|\underset{\sim}{B}^{-1}\| < 2$,

c) For all $\underset{\sim}{x} \in S_{3\eta}(\underset{\sim}{x}_1)$, $\|\underset{\sim}{J}_{\underset{\sim}{f}}(\underset{\sim}{x})-\underset{\sim}{B}\| < 1/(2\|\underset{\sim}{B}^{-1}\|)$.

Then there exists $\hat{\underset{\sim}{x}}$, a solution of $\underset{\sim}{G}(\underset{\sim}{x}) = \underset{\sim}{0}$ such that $\hat{\underset{\sim}{x}} \in S_{3\eta}(\underset{\sim}{x}_0)$.

(Note that Condition 3.2.1.v is equivalent to stating that for all $\underset{\sim}{x} \in S_{4\eta}(\underset{\sim}{x}_0)$, $\left.\frac{\partial G_i(\underset{\sim}{x})}{\partial x_j}\right|_{\underset{\sim}{x}}$ exists and is continuous, i,j=1,2,...,p.)

PROOF.

First note that 3.2.1.v implies that $\underset{\sim}{f}(\underset{\sim}{x})$ is continuously differentiable in $S_{4\eta}(\underset{\sim}{x}_0)$. Second, since $\|\underset{\sim}{x}_1 - \underset{\sim}{x}_0\| = \|\underset{\sim}{A}^{-1}\underset{\sim}{a}\| < \eta$ by 3.2.1.iii, $S_{3\eta}(\underset{\sim}{x}_1) \subset S_{4\eta}(\underset{\sim}{x}_0)$. Now proceed as in Rudin's proof. Let

$\lambda = \dfrac{1}{4\|\underset{\sim}{B}^{-1}\|}$ where $\underset{\sim}{B}$ is as above.

Then $\|\underset{\sim}{J}_{\underset{\sim}{f}}(\underset{\sim}{x}) - \underset{\sim}{B}\| < 2\lambda$ for all $\underset{\sim}{x} \in S_{3\eta}(\underset{\sim}{x}_1)$ by 3.2.1.vi.c. Now suppose $\underset{\sim}{x} \in S_{3\eta}(\underset{\sim}{x}_1)$ and $\underset{\sim}{h}$ is such that $\underset{\sim}{x} + \underset{\sim}{h} \in S_{3\eta}(\underset{\sim}{x}_1)$. Let $\underset{\sim}{F}(t) \equiv \underset{\sim}{f}(\underset{\sim}{x} + t\underset{\sim}{h}) - t\underset{\sim}{B}\underset{\sim}{h}$, $0 \le t \le 1$. (i.e. $F: [0,1] \to R^p$.) Since any norm is a convex function on R^p, $\underset{\sim}{x} + t\underset{\sim}{h} \in S_{3\eta}(\underset{\sim}{x}_1)$ for $t \in [0,1]$. Thus

$$\|\underset{\sim}{F}'(t)\| = \|\underset{\sim}{J}_{\underset{\sim}{f}}(\underset{\sim}{x} + t\underset{\sim}{h})\underset{\sim}{h} - \underset{\sim}{B}\underset{\sim}{h}\|$$

$$\le \|\underset{\sim}{J}_{\underset{\sim}{f}}(\underset{\sim}{x} + t\underset{\sim}{h}) - \underset{\sim}{B}\| \, \|\underset{\sim}{h}\|$$

by definition of $\|\cdot\|$ for matrices,

$$\leq 2\lambda \, \|\underset{\sim}{h}\|$$

by 3.2.1.vi.c,

$$= 2\lambda \, \|\underset{\sim}{B}^{-1}\underset{\sim}{Bh}\|$$

$$\leq 2\lambda \, \|\underset{\sim}{B}^{-1}\| \, \|\underset{\sim}{Bh}\|$$

$$= \tfrac{1}{2} \, \|\underset{\sim}{Bh}\|$$

by definition of λ. Theorem 5.20 of Rudin (1964:99) states: If $\underset{\sim}{F}$ is a continuous mapping of [a,b] into R^p and $\underset{\sim}{F}$ is differentiable in (a,b), then there exists $r \in (a,b)$ such that $\|\underset{\sim}{F}(b) - \underset{\sim}{F}(a)\| \leq (b-a)\,\|\underset{\sim}{F}'(r)\|$. Since $\underset{\sim}{F}$ above is indeed continuous on (0,1) this theorem applies and

$$\|\underset{\sim}{f}(\underset{\sim}{x}+\underset{\sim}{h}) - \underset{\sim}{f}(\underset{\sim}{x}) - \underset{\sim}{Bh}\| = \|\underset{\sim}{F}(1) - \underset{\sim}{F}(0)\|$$

$$\leq \|\underset{\sim}{F}'(r)\|$$

$$\leq \tfrac{1}{2} \, \|\underset{\sim}{Bh}\| \, .$$

The triangle inequality, $\|\underset{\sim}{y}+\underset{\sim}{z}\| \leq \|\underset{\sim}{y}\| + \|\underset{\sim}{z}\|$, can be applied to $\underset{\sim}{z} = \underset{\sim}{w}-\underset{\sim}{y}$ to give $\|\underset{\sim}{w}\| - \|\underset{\sim}{y}\| \leq \|\underset{\sim}{w}-\underset{\sim}{y}\| = \|\underset{\sim}{y}-\underset{\sim}{w}\|$. This property is used with $\underset{\sim}{w} = \underset{\sim}{Bh}$, $\underset{\sim}{y} = \underset{\sim}{f}(\underset{\sim}{x}+\underset{\sim}{h}) - \underset{\sim}{f}(\underset{\sim}{x})$ to give

$$\|\underset{\sim}{Bh}\| - \|\underset{\sim}{f}(\underset{\sim}{x}+\underset{\sim}{h})-\underset{\sim}{f}(\underset{\sim}{x})\| \leq \|\underset{\sim}{f}(\underset{\sim}{x}+\underset{\sim}{h})-\underset{\sim}{f}(\underset{\sim}{x})-\underset{\sim}{Bh}\|$$

$$\leq \tfrac{1}{2} \, \|\underset{\sim}{Bh}\| \, .$$

This implies

$$\|\underset{\sim}{f}(\underset{\sim}{x}+\underset{\sim}{h}) - \underset{\sim}{f}(\underset{\sim}{x})\| \geq \tfrac{1}{2}\,\|\underset{\sim}{B}\underset{\sim}{h}\|$$
$$\geq 2\lambda\,\|\underset{\sim}{h}\| \qquad (*)$$

whenever $\underset{\sim}{x}$ and $\underset{\sim}{x}+\underset{\sim}{h}$ belong to $S_{3\eta}(\underset{\sim}{x}_1)$. Hence $\underset{\sim}{f}(\cdot)$ is 1-1 on $S_{3\eta}(\underset{\sim}{x}_1)$.

Now $S_{2\eta}(\underset{\sim}{x}_1) \subset S_{3\eta}(\underset{\sim}{x}_1)$ and $\overline{S}_{2\eta}(\underset{\sim}{x}_1) \subset S_{3\eta}(\underset{\sim}{x}_1)$ ($\overline{S}$ = closure of S). To prove that $\underset{\sim}{f}[S_{2\eta}(\underset{\sim}{x}_1)]$ contains the sphere $S_{2\lambda\eta}[\underset{\sim}{f}(\underset{\sim}{x}_1)]$, note that if

$$\underset{\sim}{y}_1 = \underset{\sim}{f}(\underset{\sim}{x}_1)$$
$$= \underset{\sim}{A}^{-1}\underset{\sim}{a} + \underset{\sim}{x}_0 - \underset{\sim}{A}^{-1}\underset{\sim}{a} - \underset{\sim}{x}_0 + \underset{\sim}{A}^{-1}\underset{\sim}{r}(\underset{\sim}{x}_1)$$
$$= \underset{\sim}{A}^{-1}\underset{\sim}{r}(\underset{\sim}{x}_1),$$

then $\|\underset{\sim}{y}_1\| < \frac{\eta}{4}$ by 3.2.1.iv and $2\lambda\eta = \dfrac{2\eta}{4\|\underset{\sim}{B}^{-1}\|} > \dfrac{\eta}{4}$ because $\|\underset{\sim}{B}^{-1}\| < 2$ by 3.2.1.vi.b. Thus $\|\underset{\sim}{y}_1\| < 2\lambda\eta$ and $S_{2\lambda\eta}(\underset{\sim}{y}_1)$ contains $\underset{\sim}{0}$.

Let $T = \{\underset{\sim}{x} \mid \|\underset{\sim}{x}-\underset{\sim}{x}_1\| = 2\eta\}$. Now fix $\underset{\sim}{y} \in S_{2\lambda\eta}(\underset{\sim}{y}_1)$ (i.e. $\|\underset{\sim}{y}-\underset{\sim}{y}_1\| < 2\lambda\eta$) and define $\phi(\underset{\sim}{x}) = \|\underset{\sim}{y}-\underset{\sim}{f}(\underset{\sim}{x})\|$. It must be shown that there exists $\underset{\sim}{x}^* \in S_{2\eta}(\underset{\sim}{x}_1)$ such that $\underset{\sim}{f}(\underset{\sim}{x}^*) = \underset{\sim}{y}$. (i.e. such that $\phi(\underset{\sim}{x}^*) = 0$.)

First note that if $\underset{\sim}{x} \in T$, then (*) implies

$$4\lambda\eta = 2\lambda\|\underset{\sim}{x}-\underset{\sim}{x}_1\|$$

by definition of T,

$$\leq \|\underset{\sim}{f}[\underset{\sim}{x}_1 + (\underset{\sim}{x}-\underset{\sim}{x}_1)] - \underset{\sim}{f}(\underset{\sim}{x}_1)\|$$

by (*) with $\underset{\sim}{h} = \underset{\sim}{x}-\underset{\sim}{x}_1$,

$$\leq \|\underset{\sim}{f}(\underset{\sim}{x})-\underset{\sim}{y}\| + \|\underset{\sim}{y}-\underset{\sim}{y}_1\|$$

by triangle inequality $[\underset{\sim}{y}_1 = \underset{\sim}{f}(\underset{\sim}{x}_1)]$,

$$= \phi(\underset{\sim}{x}) + \phi(\underset{\sim}{x}_1)$$

$$< \phi(\underset{\sim}{x}) + 2\lambda\eta$$

because $\phi(x_1) = \|\underset{\sim}{y}-\underset{\sim}{y}_1\| < 2\lambda\eta$ because $\underset{\sim}{y} \in S_{2\lambda\eta}(\underset{\sim}{y}_1)$. Thus $\phi(\underset{\sim}{x}_1) < 2\lambda\eta < \phi(\underset{\sim}{x})$ for all $\underset{\sim}{x} \in T$.

Norms are continuous functions of the components and $\underset{\sim}{f}$ is continuous so $\phi(\underset{\sim}{x})$ is continuous. Further, $\bar{S}_{2\eta}(\underset{\sim}{x}_1)$ is compact. Hence there exists $\underset{\sim}{x}^* \in \bar{S}_{2\eta}(\underset{\sim}{x}_1)$ such that $\phi(\underset{\sim}{x}^*) \leq \phi(\underset{\sim}{x})$ for all $\underset{\sim}{x} \in \bar{S}_{2\eta}(\underset{\sim}{x}_1)$. But $\underset{\sim}{x}^*$ cannot belong to T because $\phi(\underset{\sim}{x}_1) < \phi(\underset{\sim}{x})$ for all $\underset{\sim}{x} \in T$. Let $\underset{\sim}{w} = \underset{\sim}{y}-\underset{\sim}{f}(\underset{\sim}{x}^*)$. Since $\underset{\sim}{B}$ is nonsingular, let $\underset{\sim}{h} = \underset{\sim}{B}^{-1}\underset{\sim}{w}$. Then choose $t \in (0,1)$ so small that $\underset{\sim}{x}^* + t\underset{\sim}{h} \in S_{2\eta}(\underset{\sim}{x}_1)$. Then

$$\begin{aligned}\|\underset{\sim}{y}-\underset{\sim}{f}(\underset{\sim}{x}^*)- \underset{\sim}{B}t\underset{\sim}{h}\| &= \|\underset{\sim}{w}-t\underset{\sim}{w}\| \\ &= \|(1-t)\underset{\sim}{w}\| \\ &= (1-t)\|\underset{\sim}{w}\|.\end{aligned}$$

But

$$\|\underset{\sim}{f}(\underset{\sim}{x}^* + t\underset{\sim}{h}) - \underset{\sim}{f}(\underset{\sim}{x}^*) - \underset{\sim}{B}t\underset{\sim}{h}\| \leq \tfrac{1}{2}\|\underset{\sim}{B}t\underset{\sim}{h}\|$$

by the argument above,

$$\begin{aligned}&= \tfrac{1}{2}\|t\underset{\sim}{w}\| \\ &= \tfrac{1}{2}t\|\underset{\sim}{w}\|.\end{aligned}$$

Then

$$\phi(\underset{\sim}{x}^* + t\underset{\sim}{h}) = \|\underset{\sim}{y}-\underset{\sim}{f}(\underset{\sim}{x}^* + t\underset{\sim}{h})\|$$

$$= \|\underset{\sim}{y}-\underset{\sim}{f}(\underset{\sim}{x}^*)-\underset{\sim}{B}t\underset{\sim}{h} + \underset{\sim}{B}t\underset{\sim}{h} + \underset{\sim}{f}(\underset{\sim}{x}^*) - \underset{\sim}{f}(\underset{\sim}{x}^* + t\underset{\sim}{h})\|$$

$$\leq \|\underset{\sim}{y}-\underset{\sim}{f}(\underset{\sim}{x}^*)-\underset{\sim}{B}t\underset{\sim}{h}\| + \|\underset{\sim}{B}t\underset{\sim}{h} + \underset{\sim}{f}(\underset{\sim}{x}^*) - \underset{\sim}{f}(\underset{\sim}{x}^* + t\underset{\sim}{h})\|$$

$$\leq (1-t)\ \|\underset{\sim}{w}\| + \tfrac{1}{2}t\ \|\underset{\sim}{w}\|$$

$$= (1-\tfrac{1}{2}t)\ \|\underset{\sim}{w}\|$$

$$= (1-\tfrac{1}{2}t)\ \phi\ (\underset{\sim}{x}^*)$$

If $\phi(x^*) > 0$, then $0 < t < 1$ implies $(1-\frac{1}{2}t) < 1$ which in turn implies $\phi(x^* + th) < \phi(x^*)$, which contradicts the minimal property of $\underset{\sim}{x}^*$. Therefore $\phi(\underset{\sim}{x}^*) = \|\underset{\sim}{y}-\underset{\sim}{f}(\underset{\sim}{x}^*)\| = 0$, which means $\underset{\sim}{y}=\underset{\sim}{f}(\underset{\sim}{x}^*)$. This argument can be used for each $\underset{\sim}{y} \in S_{2\lambda\eta}(\underset{\sim}{y}_1)$, yielding $\underset{\sim}{x}^* \in S_{2\eta}(\underset{\sim}{x}_1)$ such that $\underset{\sim}{y}=\underset{\sim}{f}(\underset{\sim}{x}^*)$. In particular it can be used for $\underset{\sim}{0} \in S_{2\lambda\eta}(\underset{\sim}{y}_1)$, yielding $\underset{\sim}{\hat{x}}$ such that $\underset{\sim}{f}(\underset{\sim}{\hat{x}}) = \underset{\sim}{0}$. But this says $\underset{\sim}{f}(\underset{\sim}{\hat{x}}) = \underset{\sim}{A}^{-1}\underset{\sim}{G}(\underset{\sim}{\hat{x}}) = \underset{\sim}{0}$ which implies $\underset{\sim}{G}(\underset{\sim}{\hat{x}}) = \underset{\sim}{0}$; furthermore $\underset{\sim}{\hat{x}} \in S_{2\eta}(\underset{\sim}{x}_1) \subset S_{3\eta}(\underset{\sim}{x}_0)$. This completes the proof of Theorem 3.2.1. |||

At this point it can be mentioned that not only does the above proof imply that the described $\underset{\sim}{\hat{x}}$ exists but that it is unique in $S_{3\eta}(\underset{\sim}{x}_0)$. Unfortunately the theorem does not yield complete uniqueness, just uniqueness in the neighborhood. In the subsequent statistical applications of this theorem, the limited uniqueness is of little value. This is why it has not been stated as a conclusion of the theorem.

3.3. A General Asymptotic Theorem

The theorem in this section is a general theorem on asymptotic results. It concerns a sequence of estimates $\{\hat{\psi}_n(Y_n)\}$ of a sequence of parameters $\{\psi_{0n}\}$. It states that for n large enough, there will be a root $\hat{\psi}_n(Y_n)$ of $G_n(\psi_n, Y_n) = 0$ near ψ_{0n} with large probability; it also states that $\hat{\psi}_n(Y_n) - \psi_{0n}$ converges in distribution to a multivariate normal distribution. In the application of this theorem, the estimates and parameters $\hat{\psi}_n$, ψ_n, and ψ_{0n} are obtained from the estimates and parameters in the usual problem $\hat{\theta}_n$, θ, and θ_0 by multiplication of each component of θ_n, θ, or θ_0 by the appropriate normalizing factor (See Section 4.3.). Furthermore, the function G_n represents the (normalized) likelihood equations; therefore, this theorem deals with maximum likelihood estimates which are solutions of the likelihood equations.

Theorem 3.3.1 is proved by demonstrating that for $G_n(\psi_n, Y_n)$ as defined, for n large enough, the conditions of Theorem 3.2.1 are true except for an event with small probability. The results of Theorem 3.2.1 then immediately imply the results of Theorem 3.3.1.

THEOREM 3.3.1. <u>For a sequence of values of n approaching infinity, define for each such n: Y_n an nX1 vector valued random variable, $a_n(Y_n)$ a pX1 vector function of Y_n, $A_n(Y_n)$ a pXp symmetric matrix function of Y_n, ψ_n a pX1 vector, $G_n(\psi_n, Y_n)$ and $R_n(\psi_n, Y_n)$ pX1 vector functions of ψ_n and Y_n, and J a pXp positive definite constant matrix. Suppose the following six conditions are true.</u>

i) For each $b > 0$, given $\epsilon > 0$ there exists $n_0(b,\epsilon)$ such that for all $n > n_0$

$$P\{\underset{\sim}{G}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)=\underset{\sim}{a}_n(\underset{\sim}{Y}_n)+\underset{\sim}{A}_n(\underset{\sim}{Y}_n)(\underset{\sim}{\psi}_n-\underset{\sim}{\psi}_{0n})+\underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) \text{ for all } \underset{\sim}{\psi}_n \in S_b(\underset{\sim}{\psi}_{0n})\} \geq 1-\epsilon.$$

ii) $\underset{\sim}{a}_n(\underset{\sim}{Y}_n) \xrightarrow{d} \mathcal{N}_p(\underset{\sim}{0},\underset{\sim}{J})$.

iii) $\underset{\sim}{A}_n(\underset{\sim}{Y}_n) + \underset{\sim}{J} \xrightarrow{p} \underset{\sim}{0}$.

iv) For each $b > 0$, given $\epsilon > 0$ and $\delta > 0$ there exists $n_0(b,\epsilon,\delta)$ such that for $n > n_0$

$$P\{\sup_{\underset{\sim}{\psi}_n \in S_b(\underset{\sim}{\psi}_{0n})} \|\underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\| < \delta\} \geq 1-\epsilon.$$

v) For each $b > 0$, given $\epsilon > 0$, there exists $n_0(b,\epsilon)$ such that for all $n > n_0$

$$P\{\text{the elements of } \underset{\sim}{J}_{\underset{\sim}{G}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) \text{ are continuous functions of } \underset{\sim}{\psi}_n \text{ in } S_b(\underset{\sim}{\psi}_{0n})\} \geq 1-\epsilon.$$

vi) If $\underset{\sim}{E}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) \equiv \underset{\sim}{J} + \underset{\sim}{J}_{\underset{\sim}{G}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)$, then for each $b > 0$, given $\epsilon > 0$ and $\delta > 0$, there exists $n_0(b,\epsilon,\delta)$ such that for all $n > n_0$

$$P\{\sup_{\underset{\sim}{\psi}_n \in S_b(\underset{\sim}{\psi}_{0n})} \|\underset{\sim}{E}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\| < \delta\} \geq 1-\epsilon.$$

Then it follows that given $\epsilon > 0$ there exists $b=b(\epsilon)$ such that $0 < b < \infty$ and $n_0=n_0(\epsilon)$ such that for each $n > n_0$

$$P\{\underline{\text{there exists a root }} \hat{\psi}_n(Y_n) \underline{\text{ of }} G_n(\psi_n, Y_n) = 0$$

$$\underline{\text{such that}}\ \hat{\psi}_n(Y_n) \in S_b(\psi_{0n})\} \geq 1-\epsilon.$$

Furthermore, for this root, $\hat{\psi}_n(Y_n) - \psi_{0n} \xrightarrow{d} \mathcal{N}_p(0, J^{-1})$.

PROOF.

Let $\epsilon > 0$ be given and let $\epsilon^* = \frac{\epsilon}{10}$. Define $D_n(Y_n) \equiv J + A_n(Y_n)$.

Then

$$A_n(Y_n) = -J + D_n(Y_n)$$

$$= -J[I - J^{-1} D_n(Y_n)].$$

If $\|D_n(Y_n)\|$ is small (which is true in probability by 3.3.1.iii) then $A_n^{-1}(Y_n)$ exists and is given by

$$A_n^{-1}(Y_n) = -[I - J^{-1}D_n(Y_n)]^{-1} J^{-1}.$$

Let

$$D_n^*(Y_n) \equiv A_n^{-1}(Y_n) + J^{-1}$$

$$= A_n^{-1}(Y_n)[J + A_n(Y_n)]J^{-1}$$

$$= A_n^{-1}(Y_n)D_n(Y_n)J^{-1}$$

$$= -[I - J^{-1}D_n(Y_n)]^{-1} J^{-1} D_n(Y_n)J^{-1}.$$

Then

$$\|\underset{\sim}{D}_n^*(\underset{\sim}{Y}_n)\| = \|[\underset{\sim}{I} - \underset{\sim}{J}^{-1} \underset{\sim}{D}_n(\underset{\sim}{Y}_n)]^{-1} \underset{\sim}{J}^{-1} \underset{\sim}{D}_n(\underset{\sim}{Y}_n) \underset{\sim}{J}^{-1}\|$$

$$\leq \|[\underset{\sim}{I} - \underset{\sim}{J}^{-1} \underset{\sim}{D}_n(\underset{\sim}{Y}_n)]^{-1}\| \cdot \|\underset{\sim}{J}^{-1}\| \cdot \|\underset{\sim}{D}_n(\underset{\sim}{Y}_n)\| \cdot \|\underset{\sim}{J}^{-1}\|$$

$$\leq \frac{\|\underset{\sim}{J}^{-1}\|^2 \cdot \|\underset{\sim}{D}_n(\underset{\sim}{Y}_n)\|}{1 - \|\underset{\sim}{J}^{-1}\underset{\sim}{D}_n(\underset{\sim}{Y}_n)\|} ,$$

if $\|\underset{\sim}{D}_n(\underset{\sim}{Y}_n)\|$ is small, since $\|\underset{\sim}{A}\underset{\sim}{B}\| \leq \|\underset{\sim}{A}\| \cdot \|\underset{\sim}{B}\|$ and $\|(\underset{\sim}{I} - \underset{\sim}{A})^{-1}\| \leq \frac{1}{1-\|\underset{\sim}{A}\|}$ when $\|\underset{\sim}{A}\| < 1$. Thus $\|\underset{\sim}{D}_n^*(\underset{\sim}{Y}_n)\|$ will be small whenever $\|\underset{\sim}{D}_n(\underset{\sim}{Y}_n)\|$ is small and hence 3.3.1.iii implies $\|\underset{\sim}{D}_n^*(\underset{\sim}{Y}_n)\| \xrightarrow{p} 0$ which in turn implies that $\underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n) \xrightarrow{p} -\underset{\sim}{J}^{-1}$. Therefore, the multivariate version of Slutsky's Theorem and 3.3.1.iii imply that $\underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{a}_n(\underset{\sim}{Y}_n) \xrightarrow{d} \underset{\sim}{\mathcal{N}}_p(\underset{\sim}{0},\underset{\sim}{J}^{-1})$.

Now choose η so that when $\underset{\sim}{Z} \sim \underset{\sim}{\mathcal{N}}_p(\underset{\sim}{0},\underset{\sim}{J}^{-1})$, $P\{\|\underset{\sim}{Z}\| \geq \eta\} < \frac{2}{3}\,\epsilon^*$. Now $\|\cdot\|$ is a continuous function of its elements; therefore $\underset{\sim}{Z}_n \xrightarrow{d} \underset{\sim}{Z}$ implies $\|\underset{\sim}{Z}_n\| \xrightarrow{d} \|\underset{\sim}{Z}\|$. Thus there exists n_1 such that for $n > n_1$

$$P\{\|\underset{\sim}{Z}_n\| \geq \eta\} < P\{\|\underset{\sim}{Z}\| \geq \eta\} + \frac{\epsilon^*}{3} = \epsilon^*$$

by definition of convergence in distribution. Letting $\underset{\sim}{Z}_n = \underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{a}_n(\underset{\sim}{Y}_n)$ we find that for $n > n_1$

$$P\{\|\underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{a}_n(\underset{\sim}{Y}_n)\| \geq \eta\} < \epsilon^* . \tag{1}$$

This proves Condition 3.2.1.iii except for an event of at most probability ϵ^*.

Now show that $P\{|\underset{\sim}{A}_n(\underset{\sim}{Y}_n)| = 0\} < \epsilon^*$ for n large. $\underset{\sim}{A}_n(\underset{\sim}{Y}_n) = -[\underset{\sim}{J} - \underset{\sim}{D}_n(\underset{\sim}{Y}_n)]$ and thus $|\underset{\sim}{A}_n(\underset{\sim}{Y}_n)| = (-1)^p|\underset{\sim}{J} - \underset{\sim}{D}_n(\underset{\sim}{Y}_n)|$. But $-\underset{\sim}{A}_n(\underset{\sim}{Y}_n)$ is a symmetric matrix and therefore its determinant can be shown to be nonzero by proving that it is positive definite; that is, it must be shown that $\lambda_{min}\{\underset{\sim}{J} - \underset{\sim}{D}_n(\underset{\sim}{Y}_n)\} > 0$. $\underset{\sim}{J}$ is positive definite; therefore, $\lambda_{min}(\underset{\sim}{J}) > 0$. Furthermore, it is true that $\max_{j=1,2,\ldots,p} |\lambda_j(\underset{\sim}{C}_n)| \to 0$ if and only if $[\underset{\sim}{C}_n]_{ij} \to 0$ for $i,j=1,2,\ldots,p$, if and only if $||\underset{\sim}{C}_n|| \to 0$. Therefore by assumption 3.3.1.iii there exists $n_2 \geq n_1$ such that for $n > n_2$

$$P\{\max_{j=1,2,\ldots,p} |\lambda_j\{\underset{\sim}{D}_n(\underset{\sim}{Y}_n)\}| > \tfrac{1}{2}\lambda_{min}(\underset{\sim}{J})\} < \epsilon^*,$$

which implies

$$P\{|\underset{\sim}{A}_n(\underset{\sim}{Y}_n)| = 0\} < \epsilon^* \tag{2}$$

This proves Condition 3.2.1.ii except for an event of at most probability ϵ^*

Since $||\underset{\sim}{D}_n^*(\underset{\sim}{Y}_n)|| \xrightarrow{p} 0$ there also exists $n_3 \geq n_2$ such that for $n > n_3$, $P\{||\underset{\sim}{D}_n^*(\underset{\sim}{Y}_n)|| > ||\underset{\sim}{J}^{-1}||\} < \epsilon^*$. But $||\underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)|| = ||-\underset{\sim}{J} + \underset{\sim}{D}_n^*(\underset{\sim}{Y}_n)|| \leq ||\underset{\sim}{J}|| + ||\underset{\sim}{D}_n^*(\underset{\sim}{Y}_n)||$ which implies for $n > n_3$

$$P\{||\underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)|| > 2||\underset{\sim}{J}^{-1}||\} < \epsilon^*.$$

Now apply 3.3.1.iv with $b = 4\eta$, $\epsilon = \epsilon^*$ and $\delta = \dfrac{\eta}{8||\underset{\sim}{J}^{-1}||}$. Then 3.3.1.iv implies that there exists $n_4 \geq n_3$ such that for $n > n_4$,

$$P\{\sup_{\underset{\sim}{\psi}_n \in S_{4\eta}(\underset{\sim}{\psi}_{0n})} \|\underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\| > \frac{\eta}{8\|\underset{\sim}{J}^{-1}\|}\} < \epsilon^*.$$

The last two inequalities together imply

$$P\{\sup_{\underset{\sim}{\psi}_n \in S_{4\eta}(\underset{\sim}{\psi}_{0n})} \|\underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\| > \frac{\eta}{4}\} < 2\ \epsilon^*. \tag{3}$$

This proves Condition 3.2.1.iv except for an event of at most probability 2 ϵ^*.

Now apply 3.3.1.i with $b = 4\eta$, $\epsilon = \epsilon^*$ to claim that there exists $n_5 \geq n_4$ such that for $n > n_4$

P{It is not true that

$$\underset{\sim}{G}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) = \underset{\sim}{a}_n(\underset{\sim}{Y}_n) + \underset{\sim}{A}_n(\underset{\sim}{Y}_n)(\underset{\sim}{\psi}_n - \underset{\sim}{\psi}_{0n}) + \underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\} < \epsilon^* \tag{4}$$

This proves Condition 3.2.1.i except for an event of at most probability ϵ^*.

Assumption 3.3.1.v with $b = 4\eta$ and $\epsilon = \epsilon^*$ guarantees that there exists $n_6 \geq n_5$ such that for $n > n_5$,

$$P\{\underset{\sim}{J}_{\underset{\sim}{G}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) \text{ is not continuous in } S_{4\eta}(\underset{\sim}{\psi}_{0n})\} < \epsilon^*. \tag{5}$$

This proves Condition 3.2.1.v except for an event of at most probability ϵ^*.

Now Conditions 3.2.1.vi.a-3.2.1.vi.c must be proved. If $\underset{\sim}{f}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) \equiv \underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{G}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)$, then $\underset{\sim}{J}_{\underset{\sim}{f}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) = \underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{J}_{\underset{\sim}{G}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)$.

But

$$\underset{\sim}{J}_{\underset{\sim}{G}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) = -\underset{\sim}{J} + \underset{\sim}{E}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n),$$

whence

$$\underset{\sim}{J}_{\underset{\sim}{f}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) = \underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{J}_{\underset{\sim}{G}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)$$

$$= [-\underset{\sim}{J}^{-1} + \underset{\sim}{D}_n^*(\underset{\sim}{Y}_n)][-\underset{\sim}{J} + \underset{\sim}{E}_n(\underset{\sim}{\psi}_n;\underset{\sim}{Y}_n)]$$

$$= \underset{\sim}{I}-\underset{\sim}{D}_n^*(\underset{\sim}{Y}_n)\underset{\sim}{J}-\underset{\sim}{J}^{-1}\underset{\sim}{E}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) + \underset{\sim}{D}_n^*(\underset{\sim}{Y}_n)\underset{\sim}{E}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)$$

$$\equiv \underset{\sim}{I} + \underset{\sim}{E}_n^*(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) \ .$$

Clearly $\|\underset{\sim}{E}_n^*(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\|$ will be small whenever $\|\underset{\sim}{D}_n^*(\underset{\sim}{Y}_n)\|$ and $\|\underset{\sim}{E}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\|$ are small, so that Conditions 3.3.1.iii and 3.3.1.vi with $b = 4\eta$ can be used to show that there exists $n_7 \geq n_6$ such that for $n > n_7$

$$P\{\sup_{\underset{\sim}{\psi}_n \epsilon S_{4\eta}(\underset{\sim}{\psi}_{0n})} \ \max_{j=1,2,\ldots,p} |\lambda_j\{\underset{\sim}{E}_n^*(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\}| > \tfrac{1}{2}\} < \epsilon^*.$$

Now if $\underset{\sim}{\psi}_{1n}$ is some particular point in $S_{4\eta}(\underset{\sim}{\psi}_{0n})$ and $\underset{\sim}{B}_n(\underset{\sim}{Y}_n) \equiv \underset{\sim}{J}_{\underset{\sim}{f}_n}(\underset{\sim}{\psi}_{1n},\underset{\sim}{Y}_n)$, the above probability statement implies (by the same method used to show $|\underset{\sim}{A}_n(\underset{\sim}{Y}_n)| \neq 0$ above) that for $n > n_7$

$$P\{|\underset{\sim}{B}_n(\underset{\sim}{Y}_n)| = 0\} < \epsilon^*. \tag{6}$$

This proves Condition 3.2.1.vi.a except for an event of at most probability ϵ^*.

Using the same reasoning as that used for $\underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)$ it is true that when $\|\underset{\sim}{E}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\|$ is small

$$\underset{\sim}{B}_n^{-1}(\underset{\sim}{Y}_n) = [\underset{\sim}{I} + \underset{\sim}{E}_n^* \ (\underset{\sim}{\psi}_{1n},\underset{\sim}{Y}_n)]^{-1}$$

$$\equiv \underset{\sim}{I} + \underset{\sim n}{E}^{**}(\underset{\sim 1n}{\psi}, \underset{\sim n}{Y}).$$

Thus

$$\|\underset{\sim n}{B}^{-1}(\underset{\sim n}{Y})\| = \|\underset{\sim}{I} + \underset{\sim n}{E}^{**}(\underset{\sim 1n}{\psi}, \underset{\sim n}{Y})\|$$

$$\leq \|\underset{\sim}{I}\| + \|\underset{\sim n}{E}^{**}(\underset{\sim 1n}{\psi}, \underset{\sim n}{Y})\|$$

$$\leq 1 + \|\underset{\sim n}{E}^{**}(\underset{\sim 1n}{\psi}, \underset{\sim n}{Y})\| .$$

But $\|\underset{\sim n}{E}^{**}(\underset{\sim 1n}{\psi}, \underset{\sim n}{Y})\|$ will be small whenever $\|\underset{\sim n}{E}^{*}(\underset{\sim 1n}{\psi}, \underset{\sim n}{Y})\|$ is small and 3.3.1.iii and 3.3.1.vi can be used to imply this. Thus again using 3.3.1.iii and 3.3.1.vi with $b = 4\eta$, it can be shown that there exists $n_8 \geq n_7$ such that for $n > n_8$ $P\{\|\underset{\sim n}{E}^{**}(\underset{\sim 1n}{\psi}, \underset{\sim n}{Y})\| > \frac{1}{5}\} < \epsilon^*$. This, of course, implies that

$$P\{\|\underset{\sim n}{B}^{-1}(Y_n)\| > \frac{6}{5}\} < \epsilon^* \tag{7}$$

This proves Condition 3.2.1.vi.b except for an event of at most probability ϵ^*.

If $\underset{\sim n}{\psi}$ is any point in $S_{4\eta}(\underset{\sim 0n}{\psi})$ then $\|\underset{\sim f_n}{J}(\underset{\sim n}{\psi}, \underset{\sim n}{Y}) - \underset{\sim n}{B}(\underset{\sim n}{Y})\| \leq \frac{1}{2\|\underset{\sim n}{B}^{-1}(\underset{\sim n}{Y})\|}$ if and only if $\|\underset{\sim f_n}{J}(\underset{\sim n}{\psi}, \underset{\sim n}{Y}) - \underset{\sim n}{B}(\underset{\sim n}{Y})\| \cdot \|\underset{\sim n}{B}^{-1}(\underset{\sim n}{Y})\| \leq \frac{1}{2}$.

Note that

$$\|\underset{\sim f_n}{J}(\underset{\sim n}{\psi}, \underset{\sim n}{Y}) - \underset{\sim n}{B}(\underset{\sim n}{Y})\|$$

$$= \|\underset{\sim}{I} + \underset{\sim n}{E}^{*}(\underset{\sim n}{\psi}, \underset{\sim n}{Y}) - [\underset{\sim}{I} + \underset{\sim n}{E}^{*}(\underset{\sim 1n}{\psi}, \underset{\sim n}{Y})]\|$$

$$= \|\underset{\sim}{E}_n^*(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) - \underset{\sim}{E}_n^*(\underset{\sim}{\psi}_{1n},\underset{\sim}{Y}_n)\|$$

$$\leq \|\underset{\sim}{E}_n^*(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\| + \|\underset{\sim}{E}_n^*(\underset{\sim}{\psi}_{1n},\underset{\sim}{Y}_n)\| .$$

If $\|\underset{\sim}{B}_n^{-1}(\underset{\sim}{Y}_n)\| \leq \frac{6}{5}$ and $\|\underset{\sim}{E}_n^*(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\| \leq \frac{1}{5}$ for all $\underset{\sim}{\psi}_n \in S_{4\eta}(\underset{\sim}{\psi}_{0n})$ then

$$\|\underset{\sim}{J}_{\underset{\sim}{f}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) - \underset{\sim}{B}_n(\underset{\sim}{Y}_n)\| \cdot \|\underset{\sim}{B}_n^{-1}(\underset{\sim}{Y}_n)\|$$

$$\leq [\|\underset{\sim}{E}_n^*(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\| + \|\underset{\sim}{E}_n^*(\underset{\sim}{\psi}_{1n},\underset{\sim}{Y}_n)\|] \cdot \|\underset{\sim}{B}_n^{-1}(\underset{\sim}{Y}_n)\|$$

$$\leq (\frac{1}{5} + \frac{1}{5}) \cdot \frac{6}{5} = \frac{12}{25} < \frac{1}{2}.$$

Again using 3.3.1.iii and 3.3.1.vi, there exists $n_9 \geq n_8$ such that for $n > n_9$

$$P\{\sup_{\underset{\sim}{\psi}_n \in S_{4\eta}(\underset{\sim}{\psi}_{0n})} \|\underset{\sim}{E}_n^*(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\| > \frac{1}{5}\} < \epsilon^*.$$

It follows that

$$P\{\sup_{\underset{\sim}{\psi}_n \in S_{4\eta}(\underset{\sim}{\psi}_{0n})} \|\underset{\sim}{J}_{\underset{\sim}{f}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) - \underset{\sim}{B}_n(\underset{\sim}{Y}_n)\| > \frac{1}{2\|\underset{\sim}{B}_n^{-1}(\underset{\sim}{Y}_n)\|}\}$$

$$\leq P\{\sup_{\underset{\sim}{\psi}_n \in S_{4\eta}(\underset{\sim}{\psi}_{0n})} \|\underset{\sim}{E}_n^*(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\| > \frac{1}{5} \text{ or } \|\underset{\sim}{B}_n^{-1}(\underset{\sim}{Y}_n)\| > \frac{6}{5}\}$$

$$\leq 2\ \epsilon^* \tag{8}$$

because $n > n_9 \geq n_8$. This proves Condition 3.2.1.vi.c except for an event of at most probability 2 ϵ^*.

It has now been shown in (1)-(8) that the conditions of Theorem 3.2.1 are true for each $n > n_9$ except for an event at most $\epsilon^*+\epsilon^*+2\epsilon^*+\epsilon^*+\epsilon^*+\epsilon^*+\epsilon^*+2\epsilon^*=10\epsilon^*=\epsilon$. Thus Theorem 3.2.1 applies and there exists $\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n)$ a solution of $\underset{\sim}{G}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) = \underset{\sim}{0}$ such that $\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n)\epsilon S_{4\eta}(\underset{\sim}{\psi}_{0n})$ with probability greater than $1-\epsilon$ for $n > n_9$.

But if $\underset{\sim}{G}_n(\underset{\sim}{\psi}_n(\underset{\sim}{Y}_n),\underset{\sim}{Y}_n) = \underset{\sim}{0}$,

$$\underset{\sim}{\psi}_n(\underset{\sim}{Y}_n)-\underset{\sim}{\psi}_{0n}= -\underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{a}_n(\underset{\sim}{Y}_n) - \underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{R}_n(\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n),\underset{\sim}{Y}_n).$$

The first term converges in distribution to $\underset{\sim}{\mathcal{N}}_p(\underset{\sim}{0},\underset{\sim}{J}^{-1})$, as was shown above. That the latter term converges in probability to $\underset{\sim}{0}$ is seen by the following remarks.

$$P\{||\underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{R}_n(\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n),\underset{\sim}{Y}_n)|| > \delta\}$$

$$\leq P\{||\underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{R}_n(\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n),\underset{\sim}{Y}_n)|| > \delta \text{ and } \hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n)\epsilon S_{4\eta}(\underset{\sim}{\psi}_{0n})\}$$

$$+ P\{\hat{\psi}_n(Y_n) \notin S_{4\eta}(\psi_{0n})\}$$

$$\leq P\{\sup_{\underset{\sim}{\psi}_n\epsilon S_{4\eta}(\underset{\sim}{\psi}_{0n})} ||\underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)|| > \delta\}$$

$$+ P\{\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n) \notin S_{4\eta}(\underset{\sim}{\psi}_{0n})\}.$$

But the first term is small for n large as shown in proving (3). The latter term is small for n large by the entire first part of the theorem.

The resulting inequality proves that $\underset{\sim}{A}_n^{-1}(\underset{\sim}{Y}_n)\underset{\sim}{R}_n(\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n),\underset{\sim}{Y}_n) \xrightarrow{p} 0$. Then the multivariate version of Slutsky's Theorem implies that $\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n) - \underset{\sim}{\psi}_{0n} \xrightarrow{d} \underset{\sim}{\mathcal{N}}_p(\underset{\sim}{0},\underset{\sim}{J}^{-1})$. Then Theorem 3.3.1 is proved with $b = 4\eta$ and $n_0 = n_9$ as above. |||

CHAPTER 4

ASYMPTOTIC THEORY FOR THE ANALYSIS OF VARIANCE MODEL

4.1. Introduction

In this chapter the maximum likelihood estimates are proved to be consistent and asymptotically normal as the size of the experimental design increases. This is done in Theorem 4.4.1, the main result of this paper. In Section 4.2 the assumptions used to prove the asymptotic properties are discussed; in Section 4.3 the setup used to prove Theorem 4.4.1 by application of Theorem 3.3.1 is explained; Theorem 4.4.1 is stated in Section 4.4 and proved in Section 4.5. (The details of the proof are given in Appendix A.) In Section 4.6 it is noted that the maximum likelihood estimates are asymptotically efficient in the sense of attaining the Cramér-Rao lower bound for the covariance matrix.

Asymptotic theory is useful as a practical tool when the experimentor has confidence that his experiment is "large enough" for the good asymptotic properties to hold. In the case of independent, identically distributed observations, one hopes one has taken enough observations; in the case of the analysis of variance, one hopes the size of the design is large enough. The device used to prove the good asymptotic properties in the analysis of variance is a "conceptual sequence of experiments." For each n of a sequence of values of n increasing to infinity, an experimental design is considered. Each experiment may be an extension of previous experiments or it may be an entirely different design. The only requirement is that the

sequence of experiments have the properties required in Section 4.2. Thus any particular experimental design encountered in practice may be thought of as a part of some sequence of designs; the experimenter hopes that the size of his particular design is "large enough" for the good asymptotic properties to hold. (In the independent, identically distributed case, there are results attempting to consider how large is "large enough"; no such attempt is made in this paper. This may be a fruitful subject for further research.) The assumptions of Section 4.2 may seem to be restrictive upon first examination; however, they rule out no experimental designs of practical interest. Thus Theorem 4.4.1 has wide applicability in the analysis of variance.

In Section 4.3 the reparameterization required for the application of Theorem 3.3.1 is discussed. It is shown that for each n the parameter vector $\underset{\sim}{\psi}_n$ is obtained from $\underset{\sim}{\theta}$ by multiplying each component of $\underset{\sim}{\theta}$ by the appropriate normalizing factor. The derivatives of the log-likelihood up to second order, which are needed for Theorem 4.4.1, are computed and a matrix used to compute the asymptotic covariance matrix is defined. Theorem 4.4.1 is stated in Section 4.4 and is proved using a sequence of lengthy lemmae, each of which proves that one or more of the conditions of Theorem 3.3.1 is true. These lemmae constitute the details of the proof and may be omitted without loss of continuity to the reader; they are presented in Section A.4.

4.2 Assumptions for Asymptotic Theory

The assumptions needed to carry out the asymptotic theory arguments will be stated and then briefly explained in this section. Under consideration will be a "conceptual sequence of experiments", each following the basic model for the analysis of variance described in Section 1.3. An experiment in this sequence may be an extension of previous experiments or an entirely different design. However, all such sequences must have the properties described by the following assumptions.

ASSUMPTION 4.2.1. n <u>and each</u> m_i, $i=1,2,\ldots,p_1$, <u>tend to infinity</u>; <u>each</u> m_i <u>can be considered a function of</u> n.

ASSUMPTION 4.2.2. <u>Let</u> $m_0 \equiv n$; <u>then for each</u> $i,j=0,1,\ldots,p_1$, <u>either</u>

$$\lim_{n\to\infty} \frac{m_i}{m_j} \equiv \rho_{ij} \quad \underline{\text{or}} \quad \lim_{n\to\infty} \frac{m_j}{m_i} \equiv \rho_{ji} \quad \underline{\text{exists}}.$$

(If $\rho_{ij} = 0$, then let $\rho_{ji} = \infty$ for notational convenience.)

Now without loss of generality, let the $\underset{\sim}{U}_i$ be labeled so that for $i < j$, $\rho_{ij} > 0$; i.e. the m_i are in decreasing order of magnitude. Generate a partition of the integers $\{0,1,\ldots,p_1\}$, $S_0, S_1, \ldots, S_c$, so that for indices i in the same set S_s, the associated m_i's have the same order of magnitude. Such a partition is generated as follows:

i) $i_0 \equiv 0$; $S_0 \equiv \{0\}$; $i_1 \equiv 1$.

ii) For $s=1,2,\ldots$, it is true that $i_s \in S_s$. Then for $i = i_s + 1, i_s + 2, \ldots$, include i in S_s until $\rho_{i_s,i} = \infty$; call the first value of i where this occurs i_{s+1}; then $i_{s+1} \in S_{s+1}$.

iii) Continue as in Step ii until p_1 has been placed in a set. Call this set S_c.

There are then c+1 sets in the partition, $S_0, S_1, \ldots, S_c$, and $S_s = \{i_s, \ldots, i_{s+1}-1\}$ (where $i_{c+1} \equiv p_1+1$ to insure S_c is correct).

Define sets S_s^* as follows:

$$S_s^* \equiv \bigcup_{t=s}^{c} S_t \quad , \quad s=1,2,\ldots,c \quad ,$$

$$S_{c+1}^* \equiv \phi \ .$$

The S_s^* are then sets of indices whose associated m_i have the same or smaller orders of magnitude when compared with m_{i_s}.

For each $i=1,2,\ldots,p_1$, $i \in S_s$ for some $s=1,2,\ldots,c$. Define sequences ν_i (depending on n) as follows:

$$\nu_i \equiv \text{rank}\,[\underset{\sim}{U}_{i_s} : \underset{\sim}{U}_{i_s+1} : \ \cdots \ : \underset{\sim}{U}_{p_1}]$$

$$- \text{rank}[\underset{\sim}{U}_{i_s} : \ \cdots \ : \underset{\sim}{U}_{i-1} : \underset{\sim}{U}_{i+1} : \ \cdots \ : \underset{\sim}{U}_{p_1}] \ ,$$

$$i=1,2,\ldots,p_1 \ ,$$

$$\nu_0 \equiv n - \text{rank}[\underset{\sim}{U}_1 : \ \cdots \ : \underset{\sim}{U}_{p_1}] \ .$$

ASSUMPTION 4.2.3. <u>Let</u> $r_i \equiv \lim_{n\to\infty} \frac{\nu_i}{m_i}$, $i=0,1,\ldots,p_1$; <u>then each of the</u> r_i <u>exists and is positive</u>.

For each $\underset{\sim}{U}_i$, $i=1,2,\ldots,p_1$, let the columns of $\underset{\sim}{U}_i$ be given by

$$\underset{\sim}{U}_i = [\underset{\sim}{u}_1^{(i)}, \underset{\sim}{u}_2^{(i)}, \ldots, \underset{\sim}{u}_{m_i}^{(i)}] \ .$$

ASSUMPTION 4.2.4. For every i and every $j \in S_s$ $j \neq i$, where $i \in S_s$, there exist two nonnegative constants, R_1 and R_2, both less than or equal to one, such that

$$\sum_{\ell=1}^{m_j} \left(\frac{\underset{\sim}{u}_\ell^{(j)\prime} \underset{\sim}{u}_k^{(i)}}{\underset{\sim}{u}_k^{(i)\prime} \underset{\sim}{u}_k^{(i)}} \right)^2 \leq R_2$$

for all but $R_1 m_i$ values of k in the set $\{1,2,\ldots,m_i\}$. Furthermore, R_1 and R_2 are such that

$$R_1 + (1-R_1)R_2 \leq \frac{1}{N(S_s) + 1} \quad ,$$

where $N(S_s)$ is the number of indices in the set S_s.

ASSUMPTION 4.2.5. Let $\underset{\sim}{\theta}_0' = (\underset{\sim}{\alpha}_0', \underset{\sim}{\sigma}_0')$, where $\underset{\sim}{\sigma}_0 = (\sigma_{00}, \sigma_{01}, \ldots, \sigma_{0p_1})'$, be the true parameter point which is being estimated and $\underset{\sim}{\Sigma}_0 \equiv \Sigma_{j=0}^{p_1} \sigma_{0j} \underset{\sim}{G}_j$ be the true covariance matrix. Then there exists a sequence ν_{p_1+1} (depending on n) and a $p_0 \times p_0$ positive definite matrix $\underset{\sim}{C}_0$ such that

$$\lim_{n\to\infty} \frac{1}{\nu_{p_1+1}} \underset{\sim}{X}' \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{X} = \underset{\sim}{C}_0 \ .$$

The following is an attempt to briefly explain these assumptions. The object of the assumptions is to rule out certain sequences of experiments for which the limiting distributions either degenerate or "blow up" (See Section 1.1.). For example, asymptotic theory requires an expanding sequence of experiments, which is what Assumption 4.2.1 requires. Assumption 4.2.2 requires that the expansion should be orderly -- sizes of various parts of the design should relate to each other in an orderly way. There is no need to consider disorganized sequences and so nothing of importance is lost by these first two assumptions.

The next three assumptions require that the sequence not be a degenerate one. That is, the matrix $\underset{\sim}{J}$ defined in Section 4.3 must not degenerate; it must be positive definite. These assumptions insure that this is so. The ν_i referred to in Assumption 4.2.3 is the dimension of the part of the linear space spanned by the columns of $\underset{\sim}{U}_i$ which is orthogonal to the space spanned by the columns of the other $\underset{\sim}{U}_j$ where $j \in S_s^*$, $j \neq i$ and $i \in S_s$. Thus ν_i is the dimension of the part of $\underset{\sim}{U}_i$ not dependent on the other $\underset{\sim}{U}_j$. Assumption 4.2.3 says that this part remains an integral part of $\underset{\sim}{U}_i$; it does not get overwhelmed by the other columns of $\underset{\sim}{U}_i$. It could be said that this assumption requires that the i^{th} effect not be "asymptotically confounded" with the effects associated with the other $\underset{\sim}{U}_j$ mentioned above. Such "asymptotically confounded" design sequences are of little interest and nothing is lost if they are ignored. It should be noted that this assumption implies that ν_i and m_i are of the same order of magnitude and hence $\nu_i \to \infty$, $i=0,1,\ldots,n$ by Assumption 4.2.1.

Assumption 4.2.4 is somewhat more difficult to explain. Its use

occurs naturally in Section A.4.1 and it is explained there also. This assumption could be described as a requirement for "almost orthogonality" of the designs (orthogonal in the language of experimental design). If in fact the designs are orthogonal in this sense, $R_1 = 0$ and $R_2 \to 0$ as $n \to \infty$. The seeming restrictiveness of this assumption is just that; it seems to be restrictive but it rules out nothing of any real interest. Any reasonable crossed or nested design sequence, whether balanced or not, will satisfy this assumption. Further light may be shed on this subject in Section 6.5. There an example is given of a design sequence for which Assumption 4.2.4 does not hold even though Assumption 4.2.3 does. (For the design sequence in Section 6.5, it can be shown that $\underset{\sim}{J}$ is not positive definite, even though Assumption 4.2.3 holds. Thus some stronger assumption like Assumption 4.2.4 is necessary in this problem.) Note that Assumption 4.2.4 requires that the columns of $\underset{\sim}{U}_i$ not be "too" dependent on the columns of the other $\underset{\sim}{U}_j$ just as Assumption 4.2.3 does, but that "too" dependent is defined in a slightly stronger sense than in Assumption 4.2.3.

Assumption 4.2.5 again rules out certain degenerate design sequences. These sequences are such that the fixed effects cannot be estimated properly. Again there is no loss in not considering such design sequences. Thus in all cases the assumptions above eliminate only disorganized or degenerate sequences of experiments which are of no real interest in any case.

The sequences $\nu_i^{\frac{1}{2}}$, $i=0,1,\ldots,p_1+1$, will become the proper normalizing sequences for the estimates of the parameters $\underset{\sim}{\alpha}$ and $\underset{\sim}{\sigma}$. $\nu_i^{\frac{1}{2}}$ is the

sequence used for σ_i, $i=0,1,\ldots,p_1$, and $\nu_{p_1+1}^{\frac{1}{2}}$ is the sequence for $\underset{\sim}{\alpha}$.

Note that only one normalizing sequence has been allowed for all the elements of $\underset{\sim}{\alpha}$. This usually is the correct thing to do; however, with a nontrivial amount of work of the same sort used in the remainder of this chapter, the theory can be extended to allow different rates for each element of $\underset{\sim}{\alpha}$.

4.3 Final Setup for Asymptotic Theory

In this section the transformation from $\underset{\sim}{\theta}$ to $\underset{\sim}{\psi}_n$ for each n is defined. The normalizing factors used will be $n_i \equiv \nu_i^{\frac{1}{2}}$, $i=0,1,\ldots,p_1+1$, where the ν_i are defined as in Section 4.2. All the ν_i and hence the n_i are considered as sequences (depending on n) increasing to infinity.

The reparameterization used will be $\underset{\sim}{\theta} \to \underset{\sim}{\psi}_n$, with $\underset{\sim}{\theta}' = (\underset{\sim}{\alpha}', \underset{\sim}{\sigma}')$ and $\underset{\sim}{\psi}_n' = (\underset{\sim}{\beta}_n', \underset{\sim}{\tau}_n')$, where $\underset{\sim}{\beta}_n = n_{p_1+1}\underset{\sim}{\alpha}$ and $[\underset{\sim}{\tau}_n]_i = n_i\sigma_i$, $i=0,1,\ldots,p_1$. The "true" parameter $\underset{\sim}{\theta}_0$ then transforms for each n to $\underset{\sim}{\psi}_{0n}' = (\underset{\sim}{\beta}_{0n}', \underset{\sim}{\tau}_{0n}')$. The log-likelihood[1] then becomes

$$\lambda_n(\underset{\sim}{\psi}_n, \underset{\sim}{Y}_n) = \log L_n(\underset{\sim}{\psi}_n, \underset{\sim}{Y}_n)$$

$$= -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\underset{\sim}{T}_n| - \frac{1}{2}\Big(\underset{\sim}{y} - \underset{\sim}{X}\frac{\underset{\sim}{\beta}_n}{n_{p_1+1}}\Big)'\underset{\sim}{T}_n^{-1}\Big(\underset{\sim}{y} - \underset{\sim}{X}\frac{\underset{\sim}{\beta}_n}{n_{p_1+1}}\Big)$$

1 The notation is abused somewhat by the use of $\lambda_n(\underset{\sim}{\psi}_n, \underset{\sim}{Y}_n)$ and later $\lambda(\underset{\sim}{y}, \underset{\sim}{\psi})$ to represent the log-likelihood function in terms of $\underset{\sim}{\psi}$ because $\lambda(\underset{\sim}{y}, \underset{\sim}{\theta})$ is also used to represent the log-likelihood function in terms of $\underset{\sim}{\theta}$ and $\lambda(\underset{\sim}{y}, \underset{\sim}{\theta}) \neq \lambda(\underset{\sim}{y}, \underset{\sim}{\psi})$ when $\underset{\sim}{\theta} = \underset{\sim}{\psi}$. However, it is clear from context which function λ is being referred to.

where $\underset{\sim}{T}_n = \sum_{i=0}^{p_1} \frac{[\underset{\sim}{\tau}_n]_i}{n_i} \underset{\sim}{G}_i$.

Now choose a norm for a $p \equiv p_0+p_1+1$ dimensional vector $\underset{\sim}{a} = (a_1,a_2,\ldots,a_p)'$ as follows: $\|\underset{\sim}{a}\| = \max_{i=1,2,\ldots,p} |a_i|$. Note that if $\hat{\underset{\sim}{\psi}}_n$ is an estimate of $\underset{\sim}{\psi}_n$ and $\hat{\underset{\sim}{\theta}}_n$ is the corresponding estimate of $\underset{\sim}{\theta}$ obtained by applying the inverse of the above transformation, then if $\|\hat{\underset{\sim}{\psi}}_n - \underset{\sim}{\psi}_{0n}\| < b$ for all $n > n_0$ then this implies that for all such n

$|[\hat{\underset{\sim}{\tau}}_n]_i - [\underset{\sim}{\tau}_{0n}]_i| = |n_i([\hat{\underset{\sim}{\sigma}}_n]_i - \sigma_{0i})| < b$, $i=0,1,\ldots,p_1$ and

$|[\hat{\underset{\sim}{\beta}}_n]_j - [\beta_{0n}]_j| = |n_{p_1+1}([\hat{\underset{\sim}{\alpha}}_n]_j - \alpha_{0j})| < b$, $j=1,2,\ldots,p_0$. This of course implies $|[\hat{\underset{\sim}{\sigma}}]_i - \sigma_{0i}| < \frac{b}{n_i}$, $i=0,1,\ldots,p_1$ and

$|[\hat{\underset{\sim}{\alpha}}_n]_j - \alpha_{0j}| < \frac{b}{n_{p_1+1}}$, $j=1,2,\ldots,p_0$; that is, the estimator $\hat{\underset{\sim}{\theta}}_n$ is approaching $\underset{\sim}{\theta}_0$, the true parameter, with each component converging at perhaps a different rate.

It is now necessary to write out the derivatives of the log likelihood up to second order to get the functions used in Theorem 4.4.1. At this point the notation of dependence on n will be suppressed. $\underset{\sim}{T}$ and $\underset{\sim}{\beta}$ will be used without the subscript n, understanding that everything--$\underset{\sim}{y}$, $\underset{\sim}{x}$ and $\underset{\sim}{G}_i$ included--depends on n. All logic is carried out for a certain value of $\underset{\sim}{y}$ which may be required to belong to a certain set (which set will have large probability).

The appropriate derivatives are given below. Matrix and vector derivatives are used where appropriate and expressions have been algebraically simplified where possible. In each case the indices i and j run from 0 to p_1.

$$\frac{\partial\lambda}{\partial\underset{\sim}{\beta}} = \frac{1}{n_{p_1+1}} \underset{\sim}{X}'\underset{\sim}{T}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\,\frac{\underset{\sim}{\beta}}{n_{p_1+1}}) ,$$

a $p_0 \times 1$ vector;

$$\frac{\partial\lambda}{\partial\tau_i} = \frac{1}{2n_i} [-\mathrm{tr}\ \underset{\sim}{T}^{-1}\underset{\sim}{G}_i+(\underset{\sim}{y}-\underset{\sim}{X}\,\frac{\underset{\sim}{\beta}}{n_{p_1+1}})'\underset{\sim}{T}^{-1}\underset{\sim}{G}_i\underset{\sim}{T}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\,\frac{\underset{\sim}{\beta}}{n_{p_1+1}})];$$

$$\frac{\partial^2\lambda}{\partial\underset{\sim}{\beta}\partial\underset{\sim}{\beta}'} = -\frac{1}{n^2_{p_1+1}} \underset{\sim}{X}'\underset{\sim}{T}^{-1}\underset{\sim}{X} ,$$

a $p_0 \times p_0$ matrix;

$$\frac{\partial^2\lambda}{\partial\tau_i\partial\underset{\sim}{\beta}} = -\frac{1}{n_i n_{p_1+1}} \underset{\sim}{X}'\underset{\sim}{T}^{-1}\underset{\sim}{G}_i\underset{\sim}{T}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\,\frac{\underset{\sim}{\beta}}{n_{p_1+1}}) ,$$

a $p_0 \times 1$ vector;

$$\frac{\partial^2\lambda}{\partial\tau_i\partial\tau_j} = \frac{1}{2n_i n_j} [\mathrm{tr}\ \underset{\sim}{T}^{-1}\underset{\sim}{G}_i\underset{\sim}{T}^{-1}\underset{\sim}{G}_j$$

$$- 2(\underset{\sim}{y}-\underset{\sim}{X}\,\frac{\underset{\sim}{\beta}}{n_{p_1+1}})'\ \underset{\sim}{T}^{-1}\underset{\sim}{G}_i\underset{\sim}{T}^{-1}\underset{\sim}{G}_j\underset{\sim}{T}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\,\frac{\underset{\sim}{\beta}}{n_{p_1+1}})] .$$

If under some conditions, the second derivatives can be shown to be continuous in a neighborhood of $\underset{\sim}{\psi}_{0n}$ (the true parameter), then under those conditions, a Taylor series can be employed, yielding

$$\left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\underset{\sim}{\psi}_i}\right|_{\underset{\sim}{\psi}_n} = \left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} + \sum_{j=1}^{p} \left.\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n^*}([\underset{\sim}{\psi}_n]_j - [\underset{\sim}{\psi}_{0n}]_j)$$

where $\underset{\sim}{\psi}_n^* = \mu\underset{\sim}{\psi}_{0n} + (1-\mu)\underset{\sim}{\psi}_n$ and $\mu \in (0,1)$. This can be written as a vector equation as follows:

$$\underset{\sim}{G}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) = \underset{\sim}{a}_n(\underset{\sim}{Y}_n) + \underset{\sim}{A}_n(\underset{\sim}{Y}_n)(\underset{\sim}{\psi}_n - \underset{\sim}{\psi}_{0n}) + \underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n),$$

where

$$[\underset{\sim}{G}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)]_i = \left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n},$$

$$[\underset{\sim}{a}_n(\underset{\sim}{Y}_n)]_i = \left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}},$$

$$[\underset{\sim}{A}_n(\underset{\sim}{Y}_n)]_{ij} = \left.\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}},$$

and

$$[\underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)]_i = \sum_{j=1}^{p}\left(\left.\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n^*} - \left.\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right)([\underset{\sim}{\psi}_n]_j - [\underset{\sim}{\psi}_{0n}]_j)$$

(Observe that $\underset{\sim}{a}_n(\underset{\sim}{Y}_n)$ and $\underset{\sim}{A}_n(\underset{\sim}{Y}_n)$ only depend on $\underset{\sim}{Y}$ since $\underset{\sim}{\psi}_{0n}$ is considered known.) For $\underset{\sim}{G}_n$ as defined above

$$[\underset{\sim}{J}_{\underset{\sim}{G}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)]_{ij} = \frac{\partial[\underset{\sim}{G}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)]_i}{\partial\psi_j}\Bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n}$$

$$= \frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\Bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n}, \quad i,j=1,2,\ldots,p.$$

All the items required to apply Theorem 3.3.1 have now been presented with the exception of the matrix $\underset{\sim}{J}$ which will be defined below. Theorem 3.3.1 will yield a sequence of estimates $\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n)$ of $\underset{\sim}{\psi}_{0n}$ which will be translated back to a sequence of estimates $\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{Y}_n)$ of $\underset{\sim}{\theta}_0$. The sequence $\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{Y}_n)$ will be a sequence of consistent and asymptotically normal roots of the likelihood equations.

Let a p×p matrix $\underset{\sim}{J}$ be defined in the following manner.

$$[\underset{\sim}{J}]_{ij} = \lim_{n\to\infty}\left[-\mathcal{E}_0\left(\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\Bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right)\right], \quad i,j=1,2,\ldots,p.$$

It is required that this matrix be positive definite. That the matrix is positive definite is proved in Section A.4.1. Some comments can be made appropriately at this point, however. Recall that $\underset{\sim}{y} \sim \eta_n(\underset{\sim}{X}\frac{\underset{\sim}{\beta}_0}{n_{p_1+1}}, \underset{\sim}{T}_0)$ when $\underset{\sim}{\psi}_n = \underset{\sim}{\psi}_{0n}$, the true parameter[2]. Further recall

2 In all subsequent writing, the dependence on n will be maintained for the vector $\underset{\sim}{\psi}$ which will be called $\underset{\sim}{\psi}_{0n}$, $\underset{\sim}{\psi}_{1n}$ or $\underset{\sim}{\psi}_{2n}$ on occasion; however the $\underset{\sim}{T}$ matrices and $\underset{\sim}{\beta}$ vectors corresponding will be called only $\underset{\sim}{T}_0$, $\underset{\sim}{T}_1$ or $\underset{\sim}{T}_2$ and $\underset{\sim}{\beta}_0$, $\underset{\sim}{\beta}_1$ or $\underset{\sim}{\beta}_2$ even though they are correctly $\underset{\sim}{T}_{0n}$, $\underset{\sim}{T}_{1n}$ etc. There is still, of course, a dependence on n even though it is suppressed in the notation.

that

$$\underset{\sim}{T}_0 = \sum_{i=0}^{p_1} \frac{[\tau_{0n}]_i}{n_i} \underset{\sim}{G}_i$$

$$= \sum_{i=0}^{p_1} \frac{n_i \sigma_{0i}}{n_i} \underset{\sim}{G}_i$$

$$= \sum_{i=0}^{p_1} \sigma_{0i} \underset{\sim}{G}_i$$

$$= \underset{\sim}{\Sigma}_0$$

and

$$\frac{\underset{\sim}{\beta}_0}{n_{p_1+1}} = \frac{n_{p_1+1} \underset{\sim}{\alpha}_0}{n_{p_1+1}}$$

$$= \underset{\sim}{\alpha}_0$$

for all values of n. Thus $\underset{\sim}{y} \sim \underset{\sim}{\eta}_n(\underset{\sim}{X}\underset{\sim}{\alpha}_0, \underset{\sim}{\Sigma}_0)$ when $\underset{\sim}{\psi} = \underset{\sim}{\psi}_{0n}$, for all values of n. Then it follows from the definitions of second derivatives given above that

$$- \mathcal{E}_0 \left(\frac{\partial^2 \lambda}{\partial \underset{\sim}{\beta} \partial \underset{\sim}{\beta}'} \bigg|_{\underset{\sim}{\psi} = \underset{\sim}{\psi}_{0n}} \right) = \frac{1}{n^2_{p_1+1}} \underset{\sim}{X}' \underset{\sim}{T}_0^{-1} \underset{\sim}{X}$$

$$= \frac{1}{n^2_{p_1+1}} \underset{\sim}{X}' \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{X}$$

$$\rightarrow \underset{\sim}{C}_0$$

by Assumption 4.2.5;

$$-\mathcal{E}_0\left(\frac{\partial^2\lambda}{\partial\tau_i\partial\underset{\sim}{\beta}}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right) = \mathcal{E}_0\left[\frac{1}{n_i n_{p_1+1}}\underset{\sim}{X}'\underset{\sim}{T}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_0^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_0}{n_{p_1+1}})\right]$$

$$= \underset{\sim}{0}\ , \quad i=0,1,\ldots,p_1\ ,$$

because $\mathcal{E}_0(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\alpha}_0$; and

$$-\mathcal{E}_0\left(\frac{\partial^2\lambda}{\partial\tau_i\partial\tau_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right) = \frac{-1}{2n_i n_j}\operatorname{tr}\underset{\sim}{T}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_0^{-1}\underset{\sim}{G}_j$$

$$+\frac{1}{n_i n_j}\mathcal{E}_0\left[(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_0}{n_{p_1+1}})'\underset{\sim}{T}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_0^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_0^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_0}{n_{p_1+1}})\right]$$

$$= \frac{1}{2n_i n_j}\operatorname{tr}\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j$$

by Lemma B.1. Thus it remains to study the properties of the $(p_1+1)\times(p_1+1)$ matrix $\underset{\sim}{C}_1$ defined by

$$(\underset{\sim}{C}_1)_{ij} = \tfrac{1}{2}\lim_{n\to\infty}\operatorname{tr}\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j,\quad i,j=0,1,\ldots,p_1.$$

It can easily be shown that the lim inf and lim sup exist for the last expression as will be done below. One of the assumptions of Theorem 4.4.1 is that the limit in fact exists -- that the lim inf equals the lim inf. This is not a grave assumption; again it merely eliminates disorganized sequences of experiments. The lim inf and lim sup can be

shown to exist if it can be proved that the sequences are bounded. Observe that

$$\operatorname{tr} \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j = \operatorname{tr} \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{U}_i\underset{\sim}{U}_i'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{U}_j\underset{\sim}{U}_j'$$

by definition of the $\underset{\sim}{G}_i$,

$$= \operatorname{tr} \underset{\sim}{U}_j'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{U}_i\underset{\sim}{U}_i'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{U}_j$$

$$= \operatorname{tr} (\underset{\sim}{U}_j'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{U}_i)(\underset{\sim}{U}_j'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{U}_i)$$

$$\geq 0$$

because for any matrix $\underset{\sim}{A}$, $\operatorname{tr} \underset{\sim}{A}\underset{\sim}{A}' = \sum_i\sum_j a_{ij}^2 \geq 0$. Furthermore

$$\frac{1}{n_i n_j}\operatorname{tr} \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j \leq \frac{\min(m_i,m_j)}{n_i n_j} \cdot \frac{1}{\sigma_{0i}\sigma_{0j}}$$

by Propositions A.3.8 and A.3.9 with $\underset{\sim}{E}_1 = \underset{\sim}{E}_2 = \underset{\sim}{\Sigma}_0^{-1}$,

$$\leq \frac{B}{\sigma_{0i}\sigma_{0j}}$$

by Proposition A.3.10, where B is a finite constant not depending on n. If B^* is defined as

$$B^* \equiv \max_{i,j=0,1,\ldots,p_1} \frac{B}{\sigma_{0i}\sigma_{0j}},$$

then B^* is finite because $\underset{\sim}{\theta}_0$ is in the interior of the parameter space and thus all the σ_{0i} are positive. Therefore

$$0 \le \frac{1}{2n_i n_j} \operatorname{tr} \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j \le \frac{1}{2} B^* \ , \quad i,j=0,1,\ldots,p_1 \ ,$$

for all n. Thus the sequences are bounded and either the limit $[\underset{\sim}{C}_1]_{ij}$ exists or the lim inf does not equal the lim sup. It remains to show that the matrix $\underset{\sim}{C}_1$ is positive definite. This is done as a part of the proof of Theorem 4.4.1 which will be stated and proved in the next sections.

4.4 Consistency and Asymptotic Normality of Maximum Likelihood Estimates

Theorem 4.4.1, the basic theorem of this paper, states that in the model of Section 1.3 and under the assumptions in Sections 1.3 and 4.2, there is a root of the likelihood equations which is consistent and asymptotically normal. Consistent estimates of the various parameters are defined to be estimates converging in probability to the true parameters. However, it has been noted that the estimates of different parameters may converge at different rates. Similarly, the estimates

are asymptotically normal, but only when normalized by the correct set of normalizing sequences, which may be different for different parameters.

This theorem will be proved by using the setup of Section 4.3 to apply Theorem 3.3.1 to this problem. The proof is given separately in Section 4.5. The details of the proof are given in Appendix A.

THEOREM 4.4.1. Consider the mixed model of the analysis of variance described in Section 1.3, under Assumptions 1.3.1-1.3.6. Consider a conceptual sequence of experiments as described in Section 4.2, under Assumptions 4.2.1-4.2.5. Let the parameter space Θ be a subset of R^p $(p=p_0+p_1+1)$ and let a typical point of that space be represented as $\underset{\sim}{\theta}' = (\underset{\sim}{\alpha}', \underset{\sim}{\sigma}')$, where $\underset{\sim}{\alpha} \in R^{p_0}$ and $\underset{\sim}{\sigma} = (\sigma_0, \sigma_1, \ldots, \sigma_{p_1})'$ has nonnegative components. Let the true parameter point $\underset{\sim}{\theta}_0$ be such that $\underset{\sim}{\sigma}_0 = (\sigma_{00}, \sigma_{01}, \ldots, \sigma_{0p_1})'$ has all positive components. Let sequences n_i, $i=0,1,\ldots,p_1+1$, increasing to infinity be defined as in Section 4.3 and assume that $\lim_{n\to\infty} \frac{1}{n_i n_j} \operatorname{tr} \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j$ exists[3], $i,j=0,.,\ldots,p_1$. Let

3 Note that this assumption only requires that the lim inf equals the lim sup for this sequence. Both of these can be proved to exist by the other assumptions. Thus this is not as grave an assumption as it might seem (See Section 4.3.).

$\underset{\sim}{J}$ be a pXp matrix defined by $\underset{\sim}{J} = \begin{bmatrix} \underset{\sim}{C}_0 & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{C}_1 \end{bmatrix}$, where $\underset{\sim}{C}_0$ is a $p_0 \times p_0$ matrix defined by $\underset{\sim}{C}_0 = \lim_{n\to\infty} \frac{1}{n^2_{p_1+1}} \underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X}$ and $\underset{\sim}{C}_1$ is a $(p_1+1)\times(p_1+1)$ matrix defined by $[\underset{\sim}{C}_1]_{ij} = \frac{1}{2} \lim_{n\to\infty} \frac{1}{n_i n_j} \operatorname{tr} \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j$, $i,j=0,1,\ldots,p_1$.

Under these conditions it follows that $\underset{\sim}{J}$ is positive definite. Furthermore, for each n there exists an estimator $\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{Y}_n) = [\hat{\underset{\sim}{\alpha}}_n'(\underset{\sim}{Y}_n), \hat{\underset{\sim}{\sigma}}'(\underset{\sim}{Y}_n)]'$ of $\underset{\sim}{\theta}$ with the following properties.

i) Given $\varepsilon > 0$ there exists $b=b(\varepsilon)$ such that $0 < b < \infty$ and $n_0=n_0(\varepsilon)$ such that for all $n > n_0$

$$P\left\{ \left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\theta_i}\right|_{\underset{\sim}{\theta}=\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{Y}_n)} = 0,\ i=1,2,\ldots,p\ ;\right.$$

$$\left|[\hat{\underset{\sim}{\alpha}}_n(\underset{\sim}{Y}_n)]_j - \alpha_{0j}\right| < \frac{b}{n_{p_1+1}},\ j=1,2,\ldots,p_0;$$

$$\left.\left|[\hat{\underset{\sim}{\sigma}}_n(\underset{\sim}{Y}_n)]_i - \sigma_{0i}\right| < \frac{b}{n_i},\ i=0,1,\ldots,p_1\right\} \geq 1-\varepsilon.$$

ii) The pX1 vector whose first p_0 components are $n_{p_1+1}\{\hat{\underset{\sim}{\alpha}}_n(\underset{\sim}{Y}_n) - \underset{\sim}{\alpha}_0\}$ and whose $(p_0+i+1)^{\text{th}}$ component is

$n_i\{[\hat{\underset{\sim}{\sigma}}_n(\underset{\sim}{Y}_n)]_i - \sigma_{0i}\}$, $i=0,1,\ldots,p_1$, <u>converges in distribution to a</u> $\underset{\sim}{\mathcal{N}}_p(\underset{\sim}{0},\underset{\sim}{J}^{-1})$ <u>random variable</u>.

The proof of Theorem 4.4.1 follows in Section 4.5.

4.5. Proof of Theorem 4.4.1--Consistency and Asymptotic Normality of Maximum Likelihood Estimates.

As suggested in Section 4.3, Theorem 4.4.1 will be proved by applying Theorem 3.3.1 to the reparameterized problem given in Section 4.3. Recall that for each n the transformation used is $\underset{\sim}{\theta} \to \underset{\sim}{\psi}_n$, where $\underset{\sim}{\theta}' = (\underset{\sim}{\alpha}', \underset{\sim}{\sigma}')$, $\underset{\sim}{\psi}_n' = (\underset{\sim}{\beta}_n', \underset{\sim}{\tau}_n')$, $\underset{\sim}{\beta}_n = n_{p_1+1}\,\underset{\sim}{\alpha}$, and $[\underset{\sim}{\tau}_n]_i = n_i\sigma_i$, $i=0,1,\ldots,p_1$. Then a Taylor Series is used to write the log-likelihood as follows:

$$\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i}\Bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n} = \frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i}\Bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} + \sum_{j=1}^{p} \frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\Bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n^*}([\underset{\sim}{\psi}_n]_j - [\underset{\sim}{\psi}_{0n}]_j)$$

where $\underset{\sim}{\psi}_n^* = \mu\underset{\sim}{\psi}_{0n} + (1-\mu)\underset{\sim}{\psi}_n$ and $\mu \in (0,1)$. As in Section 4.3, this can be written as a vector equation in proper form for the application of Theorem 3.3.1.

$$\underset{\sim}{G}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) = \underset{\sim}{a}_n(\underset{\sim}{Y}_n) + \underset{\sim}{A}_n(\underset{\sim}{Y}_n)(\underset{\sim}{\psi}_n - \underset{\sim}{\psi}_{0n}) + \underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n),$$

where

$$[\underset{\sim}{G}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)]_i = \frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n},$$

$$[\underset{\sim}{a}_n(\underset{\sim}{Y}_n)]_i = \frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}},$$

$$[\underset{\sim}{A}_n(\underset{\sim}{Y}_n)]_{ij} = \frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}},$$

$$[\underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)]_i = \sum_{j=1}^{p}\left(\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n^*} - \frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right)([\underset{\sim}{\psi}_n]_j-[\underset{\sim}{\psi}_{0n}]_j).$$

$\underset{\sim}{a}_n(\underset{\sim}{Y}_n)$ and $\underset{\sim}{A}_n(\underset{\sim}{Y}_n)$ depend only on $\underset{\sim}{y}$ because $\underset{\sim}{\psi}_{0n}$ is derived from $\underset{\sim}{\theta}_0$ which is known.

Theorem 3.3.1 states that under certain conditions (Conditions 3.3.1.i-3.3.1.vi) the following statements are true.

Given $\epsilon > 0$, there exists $b=b(\epsilon)$ such that $0 < b < \infty$ and $n_0 = n_0(\epsilon)$ such that for each $n > n_0$

$$P\left\{\text{there exists a root } \hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n) \text{ of } \underset{\sim}{G}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) = \underset{\sim}{0} \text{ such that } \hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n) \in S_b(\underset{\sim}{\psi}_{0n})\right\} \geq 1-\epsilon\ ;$$

furthermore,

$$\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n) - \underset{\sim}{\psi}_{0n} \xrightarrow{d} \mathcal{N}_p(\underset{\sim}{0},\underset{\sim}{J}^{-1})\ .$$

These statements are easily translated back to the current problem. Define the estimator $\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{Y}_n)$ as follows:

$$\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n) = \begin{pmatrix} \hat{\underset{\sim}{\beta}}_n(\underset{\sim}{Y}_n) \\ \hat{\underset{\sim}{\tau}}_n(\underset{\sim}{Y}_n) \end{pmatrix} , \quad \hat{\underset{\sim}{\theta}}_n(\underset{\sim}{Y}_n) = \begin{pmatrix} \hat{\underset{\sim}{\alpha}}_n(\underset{\sim}{Y}_n) \\ \hat{\underset{\sim}{\sigma}}_n(\underset{\sim}{Y}_n) \end{pmatrix} ,$$

where

$$\hat{\underset{\sim}{\alpha}}_n(\underset{\sim}{Y}_n) \equiv \frac{1}{n_{p_1+1}} \hat{\underset{\sim}{\beta}}_n(\underset{\sim}{Y}_n) ,$$

and

$$[\hat{\underset{\sim}{\sigma}}_n(\underset{\sim}{Y}_n)]_i \equiv \frac{1}{n_i} [\hat{\underset{\sim}{\tau}}_n(\underset{\sim}{Y}_n)]_i , \quad i=0,1,\ldots,p_1 .$$

Thus $\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{Y}_n)$ is formed by the inverse of the transformation $\underset{\sim}{\theta} \to \underset{\sim}{\psi}_n$.

Then the statement that $\underset{\sim}{G}_n(\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n),\underset{\sim}{Y}_n) = \underset{\sim}{0}$ is equivalent to the statement that

$$\left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\underset{\sim}{\beta}}\right|_{\underset{\sim}{\psi}=\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n)} = \underset{\sim}{0} ,$$

$$\left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\tau_i}\right|_{\underset{\sim}{\psi}=\hat{\underset{\sim}{\psi}}_n(Y_n)} = 0 , \quad i=0,1,\ldots,p_1 .$$

But[4]

$$\left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\underset{\sim}{\beta}}\right|_{\underset{\sim}{\psi}=\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n)} = \frac{1}{n_{p_1+1}} \left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\underset{\sim}{\alpha}}\right|_{\underset{\sim}{\theta}=\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{Y}_n)} ,$$

4 The reader should again note that λ represents different functions on the left hand and right hand sides of the next two equations. See the first footnote of this chapter.

$$\left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\tau_i}\right|_{\underset{\sim}{\psi}=\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n)} = \frac{1}{n_i}\left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\sigma_i}\right|_{\underset{\sim}{\theta}=\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{Y}_n)}, \quad i=0,1,\ldots,p_1 .$$

Thus $\underset{\sim}{G}_n(\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n), \underset{\sim}{Y}_n) = \underset{\sim}{0}$ is equivalent to $\left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\theta_i}\right|_{\underset{\sim}{\theta}=\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{Y}_n)} = 0$, $i=1,2,\ldots,p$.

Furthermore, the statement that $\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n) \in S_b(\underset{\sim}{\psi}_{0n})$ is equivalent to the statement that $|[\hat{\underset{\sim}{\beta}}_n(\underset{\sim}{Y}_n)]_j - [\underset{\sim}{\beta}_{0n}]_j| < b$, $j=1,2,\ldots,p_0$, and $|[\hat{\underset{\sim}{\tau}}_n(\underset{\sim}{Y}_n)]_i - [\underset{\sim}{\tau}_{0n}]_i| < b$, $i=0,1,\ldots,p_1$, which is in turn equivalent to $|[\hat{\underset{\sim}{\alpha}}_n(\underset{\sim}{Y}_n)]_j - \alpha_{0j}| < \frac{b}{n_{p_1+1}}$, $j=1,2,\ldots,p_0$ and $|[\hat{\underset{\sim}{\sigma}}_n(\underset{\sim}{Y}_n)]_i - \sigma_{0i}| < \frac{b}{n_i}$, $i=0,1,\ldots,p_1$. In addition, the vector $\hat{\underset{\sim}{\psi}}_n(\underset{\sim}{Y}_n) - \underset{\sim}{\psi}_{0n} = \begin{pmatrix} \hat{\underset{\sim}{\beta}}_n(\underset{\sim}{Y}_n) - \underset{\sim}{\beta}_{0n} \\ \hat{\underset{\sim}{\tau}}_n(\underset{\sim}{Y}_n) - \underset{\sim}{\tau}_{0n} \end{pmatrix}$, and

$\hat{\underset{\sim}{\beta}}_n(\underset{\sim}{Y}_n) - \underset{\sim}{\beta}_{0n} = n_{p_1+1}(\hat{\underset{\sim}{\alpha}}_n(\underset{\sim}{Y}_n) - \underset{\sim}{\alpha}_0)$, $[\hat{\underset{\sim}{\tau}}_n(\underset{\sim}{Y}_n) - \underset{\sim}{\tau}_{0n}]_i = n_i([\hat{\underset{\sim}{\sigma}}_n(\underset{\sim}{Y}_n)]_i - \underset{\sim}{\sigma}_{0i})$, $i=0,1,\ldots,p_1$. Thus the conclusions of Theorem 3.3.1 imply the conclusions of Theorem 4.4.1 for $\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{Y}_n)$ as defined above. It then remains to show that all the conditions of Theorem 3.3.1 are satisfied under the assumptions for the problem as reparameterized.

The first thing to be shown is that the matrix $\underset{\sim}{J}$ defined in Theorem 4.4.1 is positive definite so that it can be used in Theorem 3.3.1. (The existence of the limits is guaranteed by the assumptions of Theorem 4.4.1.) Recall that a matrix $\underset{\sim}{J}$ was defined in Section 4.3 by

$$[\underset{\sim}{J}]_{ij} = \lim_{n\to\infty}\left[-\mathcal{E}_0\left(\left.\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right)\right], \quad i,j=1,2,\ldots,p.$$

This matrix is easily seen to be the same $\underset{\sim}{J}$ defined in Theorem 4.4.1. This matrix is shown to be positive definite in Lemma A.4.1 in Section A.4.1. Now it must be proved that the assumptions of Theorem 4.4.1 imply the six conditions of Theorem 3.3.1. This is done in a series of four lemmae. First each condition is reduced to a simpler statement and then these statements are proved separately, each in a separate subsection. The conditions are not considered in the order i-vi but are considered in approximate order of increasing difficulty.

Condition 3.3.1.ii requires that $\underset{\sim}{a}_n(\underset{\sim}{Y}_n) \xrightarrow{d} \underset{\sim}{\eta}_p(\underset{\sim}{0},\underset{\sim}{J})$; that is,

$$\left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\underset{\sim}{\psi}}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} \xrightarrow{d} \underset{\sim}{\eta}_p(\underset{\sim}{0},\underset{\sim}{J}).$$

This is proved directly in Lemma A.4.2 in Section A.4.2. Condition 3.3.1.iii requires that if $\underset{\sim}{D}_n(\underset{\sim}{Y}_n) \equiv \underset{\sim}{J}+\underset{\sim}{A}_n(\underset{\sim}{Y}_n)$, then $\underset{\sim}{D}_n(\underset{\sim}{Y}_n) \xrightarrow{p} \underset{\sim}{0}$. It is clearly sufficient that the convergence in probability occurs for each element of $\underset{\sim}{D}_n(\underset{\sim}{Y}_n)$; that is, it is sufficient to show that $[\underset{\sim}{D}_n(\underset{\sim}{Y}_n)]_{ij} \xrightarrow{p} 0$, $i,j=1,2,\ldots,p$. But

$$\begin{aligned}
[\underset{\sim}{D}_n(\underset{\sim}{Y}_n)]_{ij} &= [\underset{\sim}{A}_n(\underset{\sim}{Y}_n)]_{ij} + [\underset{\sim}{J}]_{ij} \\
&= \left.\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\psi=\psi_{0n}} + \lim_{n\to\infty}\left[-\mathcal{E}_0\left(\left.\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\psi=\psi_{0n}}\right)\right] \\
&= \left.\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} - \mathcal{E}_0\left(\left.\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right)
\end{aligned}$$

$$+ \mathcal{S}_0\left(\frac{\partial^2\lambda(\underset{\sim}{Y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right) - \lim_{n\to\infty} \mathcal{S}_0\left(\frac{\partial^2\lambda(\underset{\sim}{Y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right).$$

Thus $[\underset{\sim}{D}_n(\underset{\sim}{Y}_n)]_{ij}$ breaks into two terms. The second obviously converges to zero by the definition of limits of real numbers. Thus it is sufficient to prove that the first term converges to zero in probability for each i and j. This is proved directly in Lemma A.4.3 in Section A.4.3.

These first two conditions are proved directly; the remaining four will be proved indirectly in the following manner. Suppose there is a sequence of events, say $\{X_n\}$; it is of interest whether $P\{X_n\} \to 1$ as $n \to \infty$. This can be shown by showing that there exists another sequence of events $\{Z_n\}$ such that $Z_n \Rightarrow X_n$ for all n and that $P\{Z_n\} \to 1$ as $n \to \infty$. The object is to prove the probabalistic statement $P\{X_n\} \to 1$. This is done in two steps--one probabalistic and one not. Proving $P\{Z_n\} \to 1$ of course involves probability concepts; however, $Z_n \Rightarrow X_n$ is not a probabalistic statement and its proof is usually algebraic.

The first condition to be treated in the above manner is Condition 3.3.1.vi. This condition requires that for each $b > 0$, given $\epsilon > 0$ and $\delta > 0$, there exists $n_0(b,\epsilon,\delta)$ such that for all $n > n_0$

$$P\left\{\sup_{\underset{\sim}{\psi}_n \in S_b(\underset{\sim}{\psi}_{0n})} \|\underset{\sim}{E}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\| < \delta\right\} \geq 1-\epsilon ,$$

where $\underset{\sim}{E}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) \equiv \underset{\sim}{J} + \underset{\sim}{J}_{\underset{\sim}{G}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)$. Again it is sufficient to prove the bound for each element of $\underset{\sim}{E}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)$. Now

$$[\underset{\sim}{E}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)]_{ij} = [\underset{\sim}{J}_{\underset{\sim}{G}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)]_{ij} + [\underset{\sim}{J}]_{ij}$$

$$= \frac{\partial^2\lambda(\underset{\sim}{Y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n} - \lim_{n\to\infty} \mathcal{E}_0\left(\frac{\partial^2\lambda(\underset{\sim}{Y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right)$$

$$= \frac{\partial^2\lambda(\underset{\sim}{Y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n} - \frac{\partial^2\lambda(\underset{\sim}{Y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}$$

$$+ \frac{\partial^2\lambda(\underset{\sim}{Y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} - \mathcal{E}_0\left(\frac{\partial^2\lambda(\underset{\sim}{Y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right)$$

$$+ \mathcal{E}_0\left(\frac{\partial^2\lambda(\underset{\sim}{Y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right) - \lim_{n\to\infty} \mathcal{E}_0\left(\frac{\partial^2\lambda(\underset{\sim}{Y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right).$$

Thus $[\underset{\sim}{E}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)]_{ij}$ breaks into three terms. The second goes to zero in probability, as was shown above, and the third goes to zero by definition of limit, as shown above. Thus it is sufficient to bound in probability the first term. This is done by using Conditions A.2.1 and A.3.1 which were shown in Sections A.2 and A.3 to occur in probability. It is shown in Lemma A.4.4 in Section A.4.4 that for b as above, if Conditions A.2.1 and A.3.1 are true then

$$\sup_{\underset{\sim}{\psi}_n \in S_b(\underset{\sim}{\psi}_{0n})} \left| \frac{\partial^2\lambda(\underset{\sim}{Y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n} - \frac{\partial^2\lambda(\underset{\sim}{Y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} \right| \to 0$$

as $n \to \infty$, $i,j=1,2,\ldots,p$. Now to get $n_0(b,\epsilon,\delta)$ for Condition 3.3.1.vi, first choose n_1 so that for $n > n_1$

$$P\left\{\left| \frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} - \mathcal{E}_0\left(\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\left.\right)\right| < \frac{\delta}{3},$$

$$\left. i,j=1,2,\ldots,p\right\} \geq 1 - \frac{\epsilon}{2},$$

(which can be done by Lemma A.4.3). Then choose $n_2 \geq n_1$ such that for $n > n_2$

$$\left|\mathcal{E}_0\left(\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\left.\right) - \lim_{n\to\infty}\mathcal{E}_0\left(\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\left.\right)\right| < \frac{\delta}{3},$$

$i,j=1,2,\ldots,p$, (which can be done by the definition of limit). Then choose $n_3 \geq n_2$ such that for $n > n_3$,

$$P\left\{\text{Conditions A.2.1 and A.3.1 are true}\right\} \geq 1 - \frac{\epsilon}{2},$$

(which can be done by Proposition A.2.1 and the definition of limit). Finally, choose $n_0(b,\epsilon,\delta) \geq n_3$ such that for $n > n_0$

$$\sup_{\underset{\sim}{\psi}_n \in S_b(\underset{\sim}{\psi}_{0n})}\left| \frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n} - \frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\left|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}\right| < \frac{\delta}{3},$$

(which can be done by Lemma A.4.4). Then this n_0 is the desired object of Condition 3.3.1.vi. This full line of reasoning will not be reproduced

for the next three conditions. Analogous arguments of this type are used for each of the three.

Condition 3.3.1.iv requires that for each b > 0, given $\epsilon > 0$ and $\delta > 0$, there exists $n_0(b,\epsilon,\delta)$ such that for $n > n_0$

$$P\left\{ \sup_{\underset{\sim}{\psi}_n \epsilon S_b(\underset{\sim}{\psi}_{0n})} \|\underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)\| < \delta \right\} \geq 1-\epsilon \ .$$

As above, each element $[\underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)]_i$ may be considered separately. Now

$$[\underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)]_i = \sum_{j=1}^{p} \left(\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n^*} - \frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} \right)([\underset{\sim}{\psi}_n]_j - [\underset{\sim}{\psi}_{0n}]_j) \ ,$$

where $\underset{\sim}{\psi}_n^* = \mu\underset{\sim}{\psi}_{0n} + (1-\mu)\ \underset{\sim}{\psi}_n$ and $\mu\ \epsilon(0,1)$. But for $\underset{\sim}{\psi}_n \ \epsilon\ S_b(\underset{\sim}{\psi}_{0n})$,

$|[\underset{\sim}{\psi}_n]_j - [\underset{\sim}{\psi}_{0n}]_j| < b$, $j=1,2,\ldots,p$; furthermore

$$\begin{aligned} \|\underset{\sim}{\psi}_n^* - \underset{\sim}{\psi}_{0n}\| &= \|(1-\mu)\underset{\sim}{\psi}_n - (1-\mu)\underset{\sim}{\psi}_{0n}\| \\ &= (1-\mu)\|\underset{\sim}{\psi}_n - \underset{\sim}{\psi}_{0n}\| \\ &\leq (1-\mu)b, \end{aligned}$$

so that $\underset{\sim}{\psi}_n^* \ \epsilon\ S_b(\underset{\sim}{\psi}_{0n})$. Therefore it is sufficient to prove that

$$\sup_{\underset{\sim}{\psi}_n \epsilon S_b(\underset{\sim}{\psi}_{0n})} \left| \frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n} - \frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} \right| \xrightarrow{p} 0 \ , \ i,j=1,2,\ldots,p.$$

But this is exactly what is proved in Lemma A.4.4 using Conditions A.2.1 and A.3.1. Thus Condition 3.3.1.iv is proved.

Condition 3.3.1.v requires that for any $b > 0$, given $\epsilon > 0$, there exists $n_0(b,\epsilon)$ such that for all $n > n_0$

$$P\left\{\text{the elements of } \underset{\sim}{J}_{\underset{\sim}{G}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) \text{ are continuous functions of } \underset{\sim}{\psi}_n \text{ in } S_b(\underset{\sim}{\psi}_{0n})\right\} \geq 1-\epsilon.$$

By the same logic as used above it is sufficient to prove that if Conditions A.2.1 and A.3.1 are true, then

$[\underset{\sim}{J}_{\underset{\sim}{G}_n}(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n)]_{ij} = \dfrac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}$ is a unformly continuous function of $\underset{\sim}{\psi}_n$ in $S_b(\underset{\sim}{\psi}_{0n})$, $i,j=1,2,\ldots,p$. This is done in Lemma A.4.5 in Section A.4.5.

Condition 3.3.1.i requires that for each $b > 0$, given $\epsilon > 0$ there exists $n_0(b,\epsilon)$ such that for all $n > n_0$

$$P\left\{\underset{\sim}{G}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) = \underset{\sim}{a}_n(\underset{\sim}{Y}_n) + \underset{\sim}{A}_n(\underset{\sim}{Y}_n)(\underset{\sim}{\psi}_n-\underset{\sim}{\psi}_{0n}) + \underset{\sim}{R}_n(\underset{\sim}{\psi}_n,\underset{\sim}{Y}_n) \text{ for all } \underset{\sim}{\psi}_n \in S_b(\underset{\sim}{\psi}_{0n})\right\} \geq 1-\epsilon;$$

that is, the Taylor Series expansion given at the beginning of this section must be valid in $S_b(\underset{\sim}{\psi}_{0n})$ with large probability. For the expansion to be valid, it is sufficient that all second derivatives are continuous functions of $\underset{\sim}{\psi}_n$ in $S_b(\underset{\sim}{\psi}_{0n})$. Again it is sufficient to show that if Conditions A.2.1 and A.3.1 are true then

$\frac{\partial^2 \lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i \partial\psi_j}$ is a uniformly continuous function of $\underset{\sim}{\psi}_n$ in $S_b(\underset{\sim}{\psi}_{0n})$,

i,j=1,2,...,p. But this is what is proved in Lemma A.4.5. Thus Condition 3.3.1.i is proved.

As was shown above, the proofs of these six conditions enable Theorem 3.3.1 to be applied, which is used to prove Theorem 4.4.1 in the manner demonstrated in the beginning of this section. The five lemmae used to prove these conditions, plus other details, are given in Appendix A. The reader may easily omit these details without loss of continuity. |||

4.6. A Note on the Asymptotic Efficiency of the Maximum Likelihood Estimates

It has been shown in the previous sections of this chapter that the maximum likelihood estimates in the mixed model of the analysis of variance are consistent and asymptotically normal. It then is of interest to know whether these estimates are asymptotically efficient in some sense. Efficiency has been defined in several different ways by various authors. The maximum likelihood estimators in this problem are efficient in the sense that the Cramér-Rao lower bound for the covariance matrix is asymptotically attained. The Cramér-Rao lower bound for the covariance matrix is the inverse of the Fisher information matrix, which is the expected value (under the true parameter) of the Hessian matrix of second derivatives. In the case of independent, identically distributed observations, this definition needs no elaboration; in the problem under study here more explanation is required.

The matrix considered in the independent, identically distributed case is derived as follows. The likelihood for n observations is just n times the likelihood for one observation; therefore, the expected values of the derivatives for n observations are n times the expected values of the derivatives for one observation. The entire matrix is then normalized by 1/n and the limit taken. This is the rigorous definition of the information matrix $\underset{\sim}{\mathcal{J}}$. That is

$$[\underset{\sim}{\mathcal{J}}]_{ij} = -\lim_{n\to\infty} \frac{1}{n} \mathcal{E}_0 \left\{ \frac{\partial \lambda_n(\underset{\sim}{y}, \underset{\sim}{\theta})}{\partial\theta_i \partial\theta_j} \bigg|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \right\}$$

$$= -\lim_{n\to\infty} \frac{1}{n} \mathcal{E}_0 \left\{ n \frac{\partial\lambda_1(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\theta_i \partial\theta_j} \bigg|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \right\}$$

$$= -\mathcal{E}_0 \left\{ \frac{\partial\lambda_1(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\theta_i \partial\theta_j} \bigg|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \right\}, \quad i,j=1,2,\ldots,p.$$

Thus although a limit of matrices is considered, what is arrived at is the information matrix for one observation.

In the problem considered here there is no one observation which does not depend on n. Thus limits analogous to that above must be considered. However each parameter may require its own normalizing sequence. For notational convenience let n_i be the correct sequence for θ_i; that is, the px1 vector whose i^{th} component is $n_i\theta_i$ has a limiting normal distribution. (Note that each sequence n_i depends on n.) This does not agree exactly with the notation of previous sections but does make the exposition in this section easier. The definition of an information matrix in this case which is analogous to the definition in the independent, identically distributed case is

$$[\underset{\sim}{\mathcal{J}}]_{ij} = -\lim_{n\to\infty} \frac{1}{n_i n_j} \mathcal{E}_0 \left\{ \frac{\partial\lambda_n(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\theta_i \partial\theta_j} \bigg|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \right\},$$

$$i,j=1,2,\ldots,p.$$

A sequence of estimates is then said to be asymptotically efficient if the estimates are consistent and asymptotically normal and if the

asymptotic covariance matrix is the inverse of the information matrix $\underset{\sim}{\mathcal{J}}$ defined above.

The estimates in the analysis of variance were shown to be consistent and asymptotically normal in Theorem 4.4.1. Furthermore, the asymptotic covariance matrix was $\underset{\sim}{J}^{-1}$, where $\underset{\sim}{J}$ was defined in Theorem 4.4.1. But the matrix $\underset{\sim}{J}$ is just exactly the information matrix described above and hence the maximum likelihood estimates in the mixed model of the analysis of variance are asymptotically efficient in the sense of attaining the Cramér-Rao lower bound for the covariance matrix.

CHAPTER 5

COMPUTATION OF THE MAXIMUM LIKELIHOOD ESTIMATES

5.1. Introduction

This chapter contains discussions of computational procedures for the calculation of the maximum likelihood estimates in the mixed model of the analysis of variance as described in Section 1.3. In Section 5.2 an equivalent form of the likelihood equations is derived. A simplification which occurs when each $\underset{\sim}{G}_i$ can be simultaneously diagonalized and also a case where explicit solutions to the likelihood equations exist are described in Section 5.3. Since explicit solutions seldom exist, an iterative procedure is proposed in Section 5.4 to solve the highly nonlinear likelihood equations. The iterative procedure proposed here was first suggested by Anderson (1971b), (1973). J. N. K. Rao (1973) has pointed out in a personal communication to Anderson that this iterative procedure is in effect the method of scoring; this is also discussed in Section 5.4. In Section 5.5 it is demonstrated that when explicit solutions exist, the iterative procedure of Section 5.4 yields those solutions in one iteration from any starting point. In Section 5.6 the problem of avoiding negative estimates is discussed.

In Sections 5.7 and 5.8 the iterative procedure of Section 5.4 is compared with a procedure suggested by Hartley and Rao (1967). Using four sample problems given in Hartley and Vaughn (1972), it was found that The Iterative Procedure was computationally more efficient than

the Hartley-Rao-Vaughn algorithm. (In the remainder of Chapter 5 the iterative procedure proposed in Section 5.4 will be referred to as "The Iterative Procedure" and this will be abbreviated TIP. The algorithm developed by Hartley, Rao and Vaughn will be referred to as the Hartley-Rao-Vaughn algorithm and this will be abbreviated as the H-R-V algorithm.) A Monte Carlo study of The Iterative Procedure revealed that it was indeed a computationally efficient algorithm. Although it has not been proved here that The Iterative Procedure always converges, the Monte Carlo results indicate that it will always converge unless the convergence is to a set of negative estimates which make $\underset{\sim}{\Sigma}(\underset{\sim}{\sigma})$ singular. These negative estimates occur only infrequently and such negative estimates are not of interest in any case. A numerical technique which was used as a modification of The Iterative Procedure made it even more computationally efficient.

5.2. The Likelihood Equations and an Equivalent Form

As shown in Section 1.3 the likelihood equations which must be solved to obtain maximum likelihood estimates of $\underset{\sim}{\alpha}$ and $\underset{\sim}{\sigma}$ are

$$[\underset{\sim}{X}'\Big(\sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j\Big)^{-1}\underset{\sim}{X}]\underset{\sim}{\alpha} = \underset{\sim}{X}'\Big(\sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j\Big)^{-1} \underset{\sim}{y}$$

and

$$\operatorname{tr}\left(\sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j\right)^{-1} \underset{\sim}{G}_i = \operatorname{tr}\left(\sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j\right)^{-1} \underset{\sim}{G}_i \left(\sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j\right)^{-1} \underset{\sim}{C}, \quad i=0,1,\ldots,p_1 \, ,$$

where

$$\underset{\sim}{C} = (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha})(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha})' \, .$$

These may be rewritten with

$$\underset{\sim}{\Sigma}(\underset{\sim}{\sigma}) = \sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j$$

as

$$[\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{X}]\underset{\sim}{\alpha} = \underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{y} \, ,$$

$$\operatorname{tr} \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{G}_i = \operatorname{tr} \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{G}_i\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{C}, \quad i=0,1,\ldots,p_1.$$

(Note that $\underset{\sim}{\Sigma}(\underset{\sim}{\sigma})$ has an inverse because $\sigma_0 > 0$, $\sigma_i \geq 0$ $i=1,2,\ldots,p_1$, implies $\underset{\sim}{\Sigma}(\underset{\sim}{\sigma})$ is positive definite and hence nonsingular.) An equivalent form of the second set of equations is obtained as follows. The following identity is true

$$\underset{\sim}{I} = \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{\Sigma}(\underset{\sim}{\sigma})$$

$$= \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}) \left[\sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j\right]$$

$$= \sum_{j=0}^{p_1} \sigma_j \, \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{G}_j \, .$$

Substitute this identity into the second set of equations yielding for a typical left hand side

$$\operatorname{tr} \Sigma^{-1}(\sigma)G_i = \operatorname{tr} \Sigma^{-1}(\sigma)G_i\, I$$

$$= \operatorname{tr} \Sigma^{-1}(\sigma)G_i\Sigma^{-1}(\sigma)\left[\sum_{j=0}^{p_1} \sigma_j G_j\right]$$

$$= \sum_{j=0}^{p_1} \sigma_j \operatorname{tr} \Sigma^{-1}(\sigma)G_i\Sigma^{-1}(\sigma)G_j$$

$$= [B(\sigma)\sigma]_i \ ,$$

where $\sigma = (\sigma_0,\sigma_1,\ldots,\sigma_{p_1})'$ is a $(p_1+1)\times 1$ vector and $B(\sigma)$ is a $(p_1+1)\times(p_1+1)$ matrix whose i,j^{th} element is given by

$$[B(\sigma)]_{ij} = \operatorname{tr} \Sigma^{-1}(\sigma)G_i\Sigma^{-1}(\sigma)G_j \quad i,j=0,1,\ldots,p_1.$$

Now define a $(p_1+1)\times 1$ vector $c(\sigma,\alpha)$ (because C depends on α) by

$$[c(\sigma,\alpha)]_i = \operatorname{tr} \Sigma^{-1}(\sigma)G_i\Sigma^{-1}(\sigma)C \ , \quad i=0,1,\ldots,p_1.$$

Then the second set of likelihood equations can be written in matrix form as

$$B(\sigma)\sigma = c(\sigma,\alpha).$$

This form is equivalent to the original form since it results merely by multiplying by an identity matrix. ($\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{\Sigma}(\underset{\sim}{\sigma}) = \underset{\sim}{I}$ for any vector $\underset{\sim}{\sigma}$.)

5.3. Simplification of the Likelihood Equations When Each $\underset{\sim}{G}_i$ Can Be Simultaneously Diagonalized

In this section some results of Anderson (1969) are restated and expanded. Suppose that there exists an nXn orthogonal matrix $\underset{\sim}{P}$ (i.e. $\underset{\sim}{P}\underset{\sim}{P}' = \underset{\sim}{P}'\underset{\sim}{P} = \underset{\sim}{I}$.) such that $\underset{\sim}{P}'\underset{\sim}{G}_i\underset{\sim}{P} = \underset{\sim}{\Lambda}_i$ where $\underset{\sim}{\Lambda}_i$ is a diagonal matrix, i=0,1,...,p_1. Then $\underset{\sim}{G}_i = \underset{\sim}{P}\underset{\sim}{\Lambda}_i\underset{\sim}{P}'$ and

$$\underset{\sim}{\Sigma}(\underset{\sim}{\sigma}) = \sum_{j=0}^{p_1} \sigma_j \underset{\sim}{G}_j$$

$$= \sum_{j=0}^{p_1} \sigma_j \underset{\sim}{P}\underset{\sim}{\Lambda}_i\underset{\sim}{P}'$$

$$= \underset{\sim}{P}\left[\sum_{j=0}^{p_1} \sigma_j \underset{\sim}{\Lambda}_j\right]\underset{\sim}{P}'$$

$$\equiv \underset{\sim}{P}\underset{\sim}{\Lambda}(\underset{\sim}{\sigma})\underset{\sim}{P}'.$$

Furthermore,

$$\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}) = \underset{\sim}{P}\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{P}'.$$

Let $\underset{\sim}{V} = \underset{\sim}{P}'\underset{\sim}{C}\underset{\sim}{P}$ and rewrite the second set of likelihood equations as follows.

For the left hand sides,

$$\begin{aligned} \text{tr } \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{G}_i &= \text{tr } \underset{\sim}{P}\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{P}'\underset{\sim}{P}\underset{\sim}{\Lambda}_i\underset{\sim}{P}' \\ &= \text{tr } \underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{P}'\underset{\sim}{P}\underset{\sim}{\Lambda}_i\underset{\sim}{P}'\underset{\sim}{P} \\ &= \text{tr } \underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{\Lambda}_i \; . \end{aligned}$$

For the right hand sides,

$$\begin{aligned} \text{tr } \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{G}_i\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{C} &= \text{tr } \underset{\sim}{P}\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{P}'\underset{\sim}{P}\underset{\sim}{\Lambda}_i\underset{\sim}{P}'\underset{\sim}{P}\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{P}'\underset{\sim}{C}, \\ &= \text{tr } \underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{\Lambda}_i\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\sigma}) \; \underset{\sim}{V} \; . \end{aligned}$$

Let the diagonal elements of $\underset{\sim}{\Lambda}_i$ be $\lambda_k^{(i)}$, $k=1,2,\ldots,n$. Then the likelihood equations become

$$\sum_{k=1}^{n} \frac{\lambda_k^{(i)}}{\sum\limits_{j=0}^{p_1} \sigma_j \lambda_k^{(j)}} = \sum_{k=1}^{n} \frac{\lambda_k^{(i)} \; v_{kk}}{\left[\sum\limits_{j=0}^{p_1} \sigma_j \lambda_k^{(j)}\right]^2} \; , \; i=0,1,\ldots,p_1 \; ,$$

where v_{kk} is the k^{th} diagonal element of $\underset{\sim}{V}$. Similarly, the equivalent

form of the likelihood equations becomes

$$\sum_{j=0}^{p_1} \left\{ \sum_{k=1}^{n} \frac{\lambda_k^{(i)}\lambda_k^{(j)}}{\left[\sum_{\ell=0}^{p_1} \sigma\, \lambda_k^{(\ell)}\right]^2} \right\} \sigma_j = \sum_{k=1}^{n} \frac{\lambda_k^{(i)}\ v_{kk}}{\left[\sum_{\ell=0}^{p_1} \sigma\, \lambda_k^{(\ell)}\right]^2}\ , \quad i=0,1,\ldots,p_1.$$

Another simplification occurs if $\underset{\sim}{\alpha}$ can be estimated independent of $\underset{\sim}{\sigma}$. This will occur if $\underset{\sim}{X}$ is of the form $\underset{\sim}{X} = \underset{\sim}{Q}\underset{\sim}{F}$ where $\underset{\sim}{Q}$ is $n\times p_0$ and consists of p_0 of the columns of $\underset{\sim}{P}$ and $\underset{\sim}{F}$ is $p_0\times p_0$ and nonsingular. Without loss of generality let $\underset{\sim}{P} = [\underset{\sim}{Q}:\underset{\sim}{\tilde{Q}}]$; that is let $\underset{\sim}{Q}$ be the first p_0 columns of $\underset{\sim}{P}$. Also let

$\underset{\sim}{\Lambda}(\underset{\sim}{\sigma}) = \begin{bmatrix} \underset{\sim}{\tilde{\Lambda}}(\sigma) & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{\tilde{\tilde{\Lambda}}}(\underset{\sim}{\sigma}) \end{bmatrix}$ be partitioned as $\underset{\sim}{P}$. Then $\underset{\sim}{P}'\underset{\sim}{Q} = \begin{bmatrix} \underset{\sim}{Q}' \\ \underset{\sim}{\tilde{Q}}' \end{bmatrix} \underset{\sim}{Q} = \begin{bmatrix} \underset{\sim}{I} \\ \underset{\sim}{0} \end{bmatrix}$ and

$$\hat{\underset{\sim}{\alpha}} = [\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{X}]^{-1}\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{y}$$

$$= [\underset{\sim}{F}'\underset{\sim}{Q}'\underset{\sim}{P}\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{P}'\underset{\sim}{Q}\underset{\sim}{F}]^{-1}\underset{\sim}{F}'\underset{\sim}{Q}'\underset{\sim}{P}\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{P}'\underset{\sim}{y}$$

$$= [\underset{\sim}{F}'[\underset{\sim}{I}\ \underset{\sim}{0}]\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\sigma})\begin{bmatrix} \underset{\sim}{I} \\ \underset{\sim}{0} \end{bmatrix}\underset{\sim}{F}]^{-1}\underset{\sim}{F}'[\underset{\sim}{I}\ \underset{\sim}{0}]\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\sigma})\begin{bmatrix} \underset{\sim}{Q}' \\ \underset{\sim}{\tilde{Q}}' \end{bmatrix}\underset{\sim}{y}$$

$$= [\underset{\sim}{F}'\underset{\sim}{\tilde{\Lambda}}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{F}]^{-1}\underset{\sim}{F}'\underset{\sim}{\tilde{\Lambda}}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{Q}'\underset{\sim}{y}$$

$$= \underset{\sim}{F}^{-1}\tilde{\underset{\sim}{\Lambda}}(\underset{\sim}{\sigma})\underset{\sim}{F}^{-t}\underset{\sim}{F}'\tilde{\underset{\sim}{\Lambda}}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{Q}'\underset{\sim}{Y}$$

$$= \underset{\sim}{F}^{-1}\underset{\sim}{Q}'\underset{\sim}{Y}$$

which is independent of $\underset{\sim}{\sigma}$.

One further simplification occurs which allows a closed form solution of the likelihood equations. Following Anderson (1969:58-59), let each

$$\underset{\sim}{\Lambda}_i = \begin{bmatrix} \mu_{1i}\underset{\sim}{I} & \underset{\sim}{0} & \cdots & \underset{\sim}{0} \\ \underset{\sim}{0} & \mu_{2i}\underset{\sim}{I} & \cdots & \underset{\sim}{0} \\ \vdots & \vdots & & \vdots \\ \underset{\sim}{0} & \underset{\sim}{0} & \cdots & \mu_{p_2 i}\underset{\sim}{I} \end{bmatrix},$$

where the orders of the identities on the main diagonal are $n_1, n_2, \ldots, n_{p_2}$ and the n_k's do not depend on i. The object is to take account of the multiplicities of the roots of $\underset{\sim}{\Sigma}(\underset{\sim}{\sigma})$ or equivalently the diagonal elements of $\underset{\sim}{\Lambda}(\underset{\sim}{\sigma})$. Define V_k for $k=1,2,\ldots,p_2$ by

$$V_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} v_{jj},$$

$$V_2 = \frac{1}{n_2}\sum_{j=n_1+1}^{n_1+n_2} v_{jj},$$

...

$$V_{p_2} = \frac{1}{n_{p_2}} \sum_{j=n-n_{p_2}+1}^{n} v_{jj},$$

where the v_{jj} are defined by $\underset{\sim}{V} = \underset{\sim}{P}'\underset{\sim}{C}\underset{\sim}{P} = \underset{\sim}{P}'(\underset{\sim}{y}-\underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})(\underset{\sim}{y}-\underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})'\underset{\sim}{P}$. The likelihood equations then simplify to

$$[\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{X}]\ \underset{\sim}{\alpha} = \underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{y}$$

and

$$\sum_{k=1}^{p_2} \frac{n_k \mu_{ki}}{\left[\sum_{\ell=0}^{p_1} \sigma_\ell \mu_{k\ell}\right]} = \sum_{k=1}^{p_2} \frac{n_k\, \mu_{ki}\, V_k}{\left[\sum_{\ell=0}^{p_1} \sigma_\ell \mu_{k\ell}\right]^2}, \quad i=0,1,\ldots,p_1.$$

If $\hat{\underset{\sim}{\alpha}}$ can be solved for independent of $\underset{\sim}{\sigma}$ and if the above notation change is made, then $\hat{\underset{\sim}{\alpha}}$, $V_1, V_2, \ldots, V_{p_2}$ are a sufficient set of statistics for the problem. Explicit solutions can be found in the case of $p_2 = p_1+1$. Then the equations

$$\sum_{j=0}^{p_1} \sigma_j\, \mu_{kj} = V_k \qquad k=1,2,\ldots,p_2$$

have p_2 equations in $p_1+1 = p_2$ unknowns and can be solved for $\underset{\sim}{\sigma}$. (As Anderson points out, the matrix of coefficients will be nonsingular

because $\underset{\sim}{G}_0, \underset{\sim}{G}_1, \ldots, \underset{\sim}{G}_{p_1}$ are linearly independent [Assumption 1.3.5], which implies that $\underset{\sim}{\Lambda}_0, \underset{\sim}{\Lambda}_1, \ldots, \underset{\sim}{\Lambda}_{p_1}$ are linearly independent.) Substituting these values of $\underset{\sim}{\sigma}$ into the denominators of the likelihood equations yields

$$\sum_{k=1}^{p_2} \frac{n_k \mu_{ki}}{V_k} = \sum_{k=1}^{p_2} \frac{n_k \mu_{ki} V_k}{V_k^2}$$

which shows that the $\underset{\sim}{\sigma}$ obtained above is indeed an explicit solution of the likelihood equations.

Explicit solutions are interesting and helpful when they exist but unfortunately they often do not exist. Thus an iterative procedure is required to solve the very nonlinear likelihood equations. Such a procedure is discussed in the next section.

5.4. The Iterative Procedure

Writing the likelihood equations in the equivalent matrix form suggests a convenient, simple iterative procedure for their solution. It is basically the method of functional iteration which is used often

in numerical analysis. The equations to solve are

$$[\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{X}]\ \underset{\sim}{\alpha} = \underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\ \underset{\sim}{y}$$

and

$$\underset{\sim}{B}(\underset{\sim}{\sigma})\ \underset{\sim}{\sigma} = \underset{\sim}{c}(\underset{\sim}{\sigma},\underset{\sim}{\alpha}).$$

Since it is assumed that $\underset{\sim}{X}$ has full rank and that $\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})$ is positive definite, $\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{X}$ is nonsingular and the equation for $\underset{\sim}{\alpha}$ can be solved to give

$$\underset{\sim}{\alpha}(\underset{\sim}{\sigma}) = [\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\underset{\sim}{X}]^{-1}\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma})\ \underset{\sim}{y}\ .$$

Then there is only one equation to solve for $\underset{\sim}{\sigma}$, namely

$$\underset{\sim}{B}(\underset{\sim}{\sigma})\ \underset{\sim}{\sigma} = \underset{\sim}{c}[\underset{\sim}{\sigma},\underset{\sim}{\alpha}(\underset{\sim}{\sigma})]\ .$$

If $\underset{\sim}{B}(\underset{\sim}{\sigma})$ is nonsingular the equation may be restated as

$$\underset{\sim}{\sigma} = \underset{\sim}{B}^{-1}(\underset{\sim}{\sigma})\ \underset{\sim}{c}[\underset{\sim}{\sigma},\underset{\sim}{\alpha}(\underset{\sim}{\sigma})]\ .$$

If a $(p_1+1)\times 1$ vector $\hat{\underset{\sim}{\sigma}}$ which satisfies the last equation can be found, then $\hat{\underset{\sim}{\sigma}}$ and $\hat{\underset{\sim}{\alpha}} = \underset{\sim}{\alpha}(\hat{\underset{\sim}{\sigma}})$ will satisfy the likelihood equations. The last equation suggests the following iterative procedure. (This iterative procedure will be referred to in this and successive sections as The Iterative Procedure, to be abbreviated TIP.)

Let $\underset{\sim}{\sigma}_{(0)}$ be any initial guess for $\underset{\sim}{\sigma}$. (The problem of choosing $\underset{\sim}{\sigma}_{(0)}$ is discussed later in this section.) Then for i=0,1,2,...

$$\underset{\sim}{\sigma}_{(i+1)} = \underset{\sim}{B}^{-1}(\underset{\sim}{\sigma}_{(i)})\ c[\underset{\sim}{\sigma}_{(i)}, \underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(i)})] .$$

This process continues until $\underset{\sim}{\sigma}_{(i+1)}$ is sufficiently close to $\underset{\sim}{\sigma}_{(i)}$ in some norm. If the iterative procedure is to make sense it must be proved that $\underset{\sim}{B}(\underset{\sim}{\sigma}_{(i)})$ is nonsingular. This is done in the next proposition, which is stated and proved as Lemma 2.1 in Anderson (1971:11-12). It is restated and reproved here for completeness.

PROPOSITION 5.4.1. <u>If $\underset{\sim}{\sigma}_{(i)}$ is such that $\Sigma(\underset{\sim}{\sigma}_{(i)})$ is nonsingular then $\underset{\sim}{B}(\underset{\sim}{\sigma}_{(i)})$ is positive definite</u>.

PROOF.

Given any $(p_1+1)\times 1$ vector $\underset{\sim}{\delta} = (\delta_0, \delta_1, \ldots, \delta_{p_1})'$, it must be shown that $\underset{\sim}{\delta}'\underset{\sim}{B}(\underset{\sim}{\sigma}_{(i)})\underset{\sim}{\delta}$ is positive.

But

$$\underset{\sim}{\delta}'\underset{\sim}{B}(\underset{\sim}{\sigma}_{(i)})\underset{\sim}{\delta} = \sum_{j=0}^{p_1} \sum_{k=0}^{p_1} [\underset{\sim}{B}(\underset{\sim}{\sigma}_{(i)})]_{jk}\delta_j\delta_k ,$$

$$= \sum_{j=0}^{p_1} \sum_{k=0}^{p_1} \operatorname{tr} \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(i)})\underset{\sim}{G}_j\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(i)})\underset{\sim}{G}_k\delta_j\delta_k ,$$

$$= \operatorname{tr} \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(i)}) \left[\sum_{j=0}^{p_1} \delta_j\underset{\sim}{G}_j\right]\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(i)}) \left[\sum_{k=0}^{p_1} \delta_k\underset{\sim}{G}_k\right] ,$$

$$= \operatorname{tr}\left[\, \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(i)}) \left(\sum_{j=0}^{p_1} \delta_j \underset{\sim}{G}_j\right)\right]^2 ,$$

which is positive unless $\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(i)})\left(\sum_{j=0}^{p_1} \delta_j \underset{\sim}{G}_j\right) = \underset{\sim}{0}$. But it is impossible for $\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(i)})\left(\sum_{j=0}^{p_1} \delta_j \underset{\sim}{G}_j\right)$ to be $\underset{\sim}{0}$ because the $\underset{\sim}{G}_j$ are linearly independent and $\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(i)})$ is obviously nonsingular. |||

It should be noted that so long as $[\underset{\sim}{\sigma}_{(i)}]_0$, the 0^{th} component of $\underset{\sim}{\sigma}_{(i)}$, is positive and the others are nonnegative $\underset{\sim}{\Sigma}(\underset{\sim}{\sigma}_{(i)})$ will always be nonsingular. Thus so long as the iterative process avoids negative values it will continue unimpeded. (For a discussion of negative values for components of $\hat{\underset{\sim}{\sigma}}$, see Section 5.6.)

J. N. K. Rao has pointed out that The Iterative Procedure is in effect the method of scoring. That this is so can be seen by the following remarks. The method of scoring obtains iterations in the following manner. Given an estimate $\underset{\sim}{\theta}_{(k)}$ of the pX1 parameter $\underset{\sim}{\theta}$, calculate $\underset{\sim}{\mathcal{I}}(\underset{\sim}{\theta}_{(k)})$, the pXp information matrix whose i,j^{th} element is given by

$$[\underset{\sim}{\mathcal{I}}(\underset{\sim}{\theta}_{(k)})]_{ij} = -\mathcal{E}\left\{ \frac{\partial^2 \lambda(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\theta_i \partial\theta_j} \Bigg|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_{(k)}} \right\} , \quad i,j=1,2,\ldots,p, \text{ and}$$

$\left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\underset{\sim}{\theta}}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_{(k)}}$, the pX1 vector whose i^{th} component is $\left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\theta_i}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_{(k)}}$.

Then the next iterate $\underset{\sim}{\theta}_{(k+1)}$ is given by

$$\underset{\sim}{\theta}_{(k+1)} = \underset{\sim}{\theta}_{(k)} + \underset{\sim}{\mathcal{I}}^{-1}(\underset{\sim}{\theta}_{(k)}) \left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\underset{\sim}{\theta}}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_{(k)}}.$$

It is easily seen that $\underset{\sim}{\mathcal{I}}(\underset{\sim}{\theta}_{(k)})$ for the case under study here is given by (The notation used here is same as that used previously in this section.)

$$\underset{\sim}{\mathcal{I}}(\underset{\sim}{\theta}_{(k)}) = \begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(k)})\underset{\sim}{X} & \underset{\sim}{0} \\ \underset{\sim}{0}' & \frac{1}{2}\underset{\sim}{B}(\underset{\sim}{\sigma}_{(k)}) \end{bmatrix}.$$

Furthermore,

$$\left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\underset{\sim}{\theta}}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_k} = \begin{bmatrix} \underset{\sim}{b}_1 \\ \underset{\sim}{b}_2 \end{bmatrix},$$

where

$$\underset{\sim}{b}_1 = \underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(k)})[\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(k)})]$$

$$[\underset{\sim}{b}_2]_i = \tfrac{1}{2}[\mathrm{tr}\ \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(k)})\underset{\sim}{G}_i\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(k)})\underset{\sim}{C} - \mathrm{tr}\ \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(k)})\underset{\sim}{G}_i]$$

$$= \tfrac{1}{2}\left\{\underset{\sim}{c}[\underset{\sim}{\sigma}_{(k)},\underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(k)})] - \underset{\sim}{B}(\underset{\sim}{\sigma}_{(k)})\ \underset{\sim}{\sigma}_{(k)}\right\}_i, \quad i=0,1,\ldots,p_1 .$$

Then $\underset{\sim}{\theta}_{(k+1)} = (\underset{\sim}{\alpha}'_{(k+1)},\underset{\sim}{\sigma}'_{(k+1)})'$ is given by

$$\underset{\sim}{\alpha}_{(k+1)} = \underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(k)}) + [\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(k)})\underset{\sim}{X}]^{-1}\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(k)})\underset{\sim}{y} - \underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(k)})$$

$$= [\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(k)})\underset{\sim}{X}]^{-1}\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(k)})\underset{\sim}{y} ,$$

and

$$\underset{\sim}{\sigma}_{(k+1)} = \underset{\sim}{\sigma}_{(k)} + \underset{\sim}{B}^{-1}(\underset{\sim}{\sigma}_{(k)})\ \underset{\sim}{c}[\underset{\sim}{\sigma}_{(k)},\underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(k)})] - \underset{\sim}{\sigma}_{(k)}$$

$$= \underset{\sim}{B}^{-1}(\underset{\sim}{\sigma}_{(k)})\ \underset{\sim}{c}[\underset{\sim}{\sigma}_{(k)},\underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(k)})] .$$

Thus the method of scoring yields the same equations as The Iterative Procedure.

The Iterative Procedure described above has the advantage of being simple, easy to describe, and easy to program. (A computer program implementing this algorithm is described in Appendix C.) Other methods for solving the likelihood equations have been developed. Hartley and J.N.K. Rao (1967) suggested solution by the method of steepest ascent. They

numerically integrated out a system of simultaneous differential equations to obtain solutions. They gave a proof of the convergence of their algorithm. Hartley and Vaughn (1972) presented a computer program written by Vaughn (1970) implementing the Hartley-Rao algorithm. This program is quite complicated; it proceeds in the following manner. A least squares approximation is obtained to the differential equations to be solved. This approximation is numerically integrated by the Runge-Kutta method until a solution is obtained. These two steps are repeated until convergence is obtained. The reason an approximation must be used is that numerical integration methods may require large numbers of iterations and a large amount of effort is required to evaluate the likelihood equations. A comparison of The Iterative Procedure and the method of Hartley, Rao, and Vaughn on some sample problems presented in Hartley and Vaughn (1972) is given in Section 5.7. It can be said at this point that The Iterative Procedure seems to be computationally much more efficient, at least in most cases.

Choosing the initial estimator $\underset{\sim}{\sigma}_{(0)}$ is of some importance in The Iterative Procedure. The closer $\underset{\sim}{\sigma}_{(0)}$ is to the true solution, the easier it will be for the algorithm to iterate to that solution. Anderson (1971b:12-13) considers the case where the mean of $\underset{\sim}{y}$ is either known or completely unspecified and there are N observations on $\underset{\sim}{y}$. He suggests $\underset{\sim}{\sigma}_{(0)}$ the solution of the equations

$$\sum_{j=0}^{p_1} \sigma_j \operatorname{tr} \underset{\sim}{A}\underset{\sim}{G}_i\underset{\sim}{A}\underset{\sim}{G}_j = \operatorname{tr} \underset{\sim}{A}\underset{\sim}{G}_i\underset{\sim}{A}\underset{\sim}{C} \ , \quad i=0,1,\ldots,p_1 \ ,$$

where $\underset{\sim}{A}$ is an arbitrary positive definite matrix and

$\underset{\sim}{C} = \frac{1}{N} \Sigma_{k=1}^{N} (\underset{\sim}{y}_k - \underset{\sim}{\mu})(\underset{\sim}{y}_k - \underset{\sim}{\mu})'$ if $\underset{\sim}{\mu}$ is known or $\underset{\sim}{C} = \frac{1}{N-1} \Sigma_{k=1}^{N} (\underset{\sim}{y}_k - \bar{\underset{\sim}{y}})(\underset{\sim}{y}_k - \bar{\underset{\sim}{y}})'$ if $\underset{\sim}{\mu}$ is completely unspecified. In both cases $\underset{\sim}{C}$ is an unbiased estimate of $\underset{\sim}{\Sigma}$ and hence $\underset{\sim}{\sigma}_{(0)}$ will be an unbiased estimate of $\underset{\sim}{\sigma}$. One choice suggested for $\underset{\sim}{A}$ is $\underset{\sim}{I}$, the identity matrix. It has been found in sample problems that the above method does not work very well in practice. It often seems to take many iterations to reach convergence from the above $\underset{\sim}{\sigma}_{(0)}$ while other types of guesses work better. Anderson mentions some other methods for the case above. For the analysis of variance the following methods may work well. The usual analysis of variance estimates, if they are available, or approximations to them can be used. A rough approximation to the sum of squares for each factor may work fairly well if that is all that is available. Prior knowledge may be put into the initial guess if desired. If no information at all is available, $\underset{\sim}{\sigma}'_{(0)} = (1,0,\ldots,0)$ may be used; this corresponds to choosing $\underset{\sim}{A} = \underset{\sim}{I}$ in the Anderson method above. The choice of the initial guess is not critical but some attempt should be made to make reasonable choices.

The advantages and use of this method have been alluded to. At this point an attempt must be made to answer the three critical questions one must ask of any numerical procedure.

1) Does it converge?
2) If it does converge, to what does it converge? Is the result a root of the likelihood equations?
3) Is the answer unique?

Unfortunately complete answers are not known to any of these questions. An answer to all three is given in the next section for a special case. To study the first two in general Monte Carlo studies were used. These results are described in Section 5.8. More complete answers to these questions may be found after further research. At this point the excellent Monte Carlo results and the fact that The Iterative Procedure is in fact the method of scoring, which has been accepted for many years, encourage the use of The Iterative Procedure even though these questions remain unanswered.

5.5. A Case Where the Iterative Procedure Gives Exact Solutions in One Iteration

In the event that the conditions described in Section 5.3 for the existence of explicit solutions of the likelihood equations are satisfied, The Iterative Procedure works very well indeed. In fact it will yield the exact solutions to the likelihood equations in one iteration starting from any initial guess. Recall that these conditions are the following.

CONDITION 5.5.1. There exists an n×n orthogonal matrix $\underset{\sim}{P}$ such that $\underset{\sim}{P}'\underset{\sim}{G}_i\underset{\sim}{P} = \underset{\sim}{\Lambda}_i \quad i=0,1,\ldots,p_1$, where

$$\underset{\sim}{\Lambda}_i = \begin{bmatrix} \mu_{1i}\,\underset{\sim}{I} & \underset{\sim}{0} & \cdots & \underset{\sim}{0} \\ \underset{\sim}{0} & \mu_{2i}\,\underset{\sim}{I} & \cdots & \underset{\sim}{0} \\ \vdots & \vdots & & \vdots \\ \underset{\sim}{0} & \underset{\sim}{0} & & \mu_{p_2}\,\underset{\sim}{I} \end{bmatrix}$$

with the orders of the identities $n_1, n_2, \ldots, n_{p_2}$ respectively. The n_k's do not depend on i.

CONDITION 5.5.2. $\underset{\sim}{X} = \underset{\sim}{Q}\underset{\sim}{F}$ where $\underset{\sim}{Q}$ consists of (without loss of generality) the first p_0 columns of $\underset{\sim}{P}$ and $\underset{\sim}{F}$ is a nonsingular $p_0 \times p_0$ matrix.

CONDITION 5.5.3. $p_2 = p_1+1$.

When these conditions are true a solution to the likelihood equations is $\hat{\underset{\sim}{\alpha}} = \underset{\sim}{F}^{-1}\underset{\sim}{Q}'\underset{\sim}{y}$ and $\hat{\underset{\sim}{\sigma}}$ the solution of

$$\sum_{j=0}^{p_1} \sigma_j\, \mu_{kj} = V_k, \quad k=1,2,\ldots,p_2 ,$$

where the V_k are defined from the matrix $\underset{\sim}{V} = \underset{\sim}{P}'(\underset{\sim}{y}-\underset{\sim}{X}\hat{\underset{\sim}{\alpha}})(\underset{\sim}{y}-\underset{\sim}{X}\hat{\underset{\sim}{\alpha}})'\underset{\sim}{P}$ as follows

$$V_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} v_{jj} ,$$

$$V_2 = \frac{1}{n_2} \sum_{j=n_1+1}^{n_1+n_2} v_{jj} \ ,$$

$$\cdots$$

$$V_{p_2} = \frac{1}{n_{p_2}} \sum_{j=n-n_{p_2}+1}^{n} v_{jj} \ .$$

The following Proposition is true.

PROPOSITION 5.5.1. <u>If Conditions 5.5.1-5.5.3 are true, and $\underset{\sim}{\sigma}_{(0)}$ is any initial guess for $\underset{\sim}{\sigma}$ such that $\underset{\sim}{\Sigma}(\underset{\sim}{\sigma}_{(0)})$ is nonsingular, then The Iterative Procedure yields</u>

$$\underset{\sim}{\sigma}_{(1)} = \underset{\sim}{B}^{-1}(\underset{\sim}{\sigma}_{(0)}) \ \underset{\sim}{c}[\underset{\sim}{\sigma}_{(0)}, \hat{\underset{\sim}{\alpha}}(\underset{\sim}{\sigma}_{(0)})]$$

<u>where $\underset{\sim}{\sigma}_{(1)}$ is the solution to</u> $\sum_{j=0}^{p_1} \sigma_j \ \mu_{kj} = V_k$, $k=1,2,\ldots,p_2$.

PROOF.

Let $\underset{\sim}{\sigma}_{(0)} = (\sigma_0^{(0)}, \ \sigma_1^{(0)}, \ldots, \sigma_{p_1}^{(0)})'$. Then the matrix $\underset{\sim}{B}(\underset{\sim}{\sigma}_{(0)})$ is given

$$[\underset{\sim}{B}(\underset{\sim}{\sigma}_{(0)})]_{ij} = \text{tr } \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(0)})\underset{\sim}{G}_i\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\sigma}_{(0)})\underset{\sim}{G}_j$$

$$= \sum_{k=1}^{p_2} \frac{n_k\, \mu_{ki}\, \mu_{kj}}{\left(\sum_{\ell=0}^{p_1} \sigma_\ell^{(0)} \mu_{k\ell}\right)^2} .$$

The vector $\underset{\sim}{c}(\underset{\sim}{\sigma}_{(0)},\hat{\underset{\sim}{\alpha}})$ is given by (Note that $\hat{\alpha}$ is independent of $\underset{\sim}{\sigma}_{(0)}$.)

$$[\underset{\sim}{c}(\underset{\sim}{\sigma}_{(0)},\hat{\underset{\sim}{\alpha}})]_i = \sum_{k=1}^{p_2} \frac{n_k\, \mu_{ki}\, V_k}{\left(\sum_{\ell=0}^{p_1} \sigma_\ell^{(0)} \mu_{k\ell}\right)^2} .$$

Thus the equation to be solved for $\underset{\sim}{\sigma}_{(1)}$ is

$$\sum_{j=0}^{p_1} \left[\sum_{k=1}^{p_2} \frac{n_k\, \mu_{ki}\, \mu_{kj}}{\left(\sum_{\ell=0}^{p_1} \sigma_\ell^{(0)} \mu_{k\ell}\right)^2}\right] \sigma_j = \sum_{k=1}^{p_2} \frac{n_k\, \mu_{ki}\, V_k}{\left(\sum_{\ell=0}^{p_1} \sigma_\ell^{(0)} \mu_{k\ell}\right)^2} .$$

Since $\underset{\sim}{\Sigma}(\underset{\sim}{\sigma}_{(0)})$ is nonsingular, $\underset{\sim}{B}(\underset{\sim}{\sigma}_{(0)})$ is positive definite and, therefore there is exactly one solution to the above equation. But $\underset{\sim}{\sigma}$, the solution of $\sum_{j=0}^{p_1} \sigma_j\, \mu_{kj} = V_k$, $k=1,2,\ldots,p_2$, is a solution as is verified by the following substitution in the left hand side of each equation.

$$\sum_{j=0}^{p_1}\left[\sum_{k=1}^{p_2}\frac{n_k\,\mu_{ki}\,\mu_{kj}}{\left(\sum_{\ell=0}^{p_1}\sigma_\ell^{(0)}\mu_{k\ell}\right)^2}\right]\sigma_j = \sum_{k=1}^{p_2}\frac{n_k\,\mu_{ki}\left(\sum_{j=0}^{p_1}\mu_{kj}\,\sigma_j\right)}{\left(\sum_{\ell=0}^{p_1}\sigma_\ell^{(0)}\,\mu_{k\ell}\right)^2}$$

$$= \sum_{k=1}^{p_2}\frac{n_k\,\mu_{ki}\,V_k}{\left(\sum_{\ell=0}^{p_1}\sigma_\ell^{(0)}\,\mu_{k\ell}\right)^2}\ ,$$

which is the right hand side of the equation. Thus the unique solution $\underset{\sim}{\sigma}_{(1)}$ is just $\hat{\underset{\sim}{\sigma}}$ the solution to

$$\sum_{j=0}^{p_1}\sigma_j\,\mu_{kj} = V_k \text{ for } k=1,2,\ldots,p_2. \quad |||$$

It is reassuring to know that if there exist exact solutions, The Iterative Procedure will pick them out immediately even if the user of the method is unaware that they exist. The Hartley-Rao-Vaughn algorithm does not have this property. However it does have the property that it contains built in traps to guarantee that it never produces negative estimates. Of course the maximum likelihood estimates are never negative. The Iterative Procedure as presently constructed does not inherently avoid negative estimates. However, it can be adjusted so that the final answers are nonnegative. This is discussed in the next section.

5.6. Avoiding Negative Estimates

The general problem of negative estimates of variance components has been discussed by many authors. (See Searle [1971:22] for a summary.) Of course, maximum likelihood estimates cannot be negative; negative estimates do not belong in the parameter space and hence are ineligible for maximum likelihood estimation.

The Iterative Procedure defined above does not by its nature converge to nonnegative estimates. In fact it may very well converge to negative estimates. However, a simple modification allows one to avoid negative estimates as final answers. The general solution when any estimates are computed as negative is to fix them at zero and solve the remaining likelihood equations subject to these constraints; that is, if σ_i would have been negative for $i \in S$, where S is some set of indices, the new equations become

$$\frac{\partial\lambda}{\partial\underset{\sim}{\alpha}} = \underset{\sim}{0} ,$$

$$\frac{\partial\lambda}{\partial\sigma_j} = 0 , \quad j=0,.,\ldots,p_1 \quad j \notin S,$$

$$\sigma_i = 0 \qquad i \in S.$$

This is the method used by Hartley, Rao, and Vaughn in their algorithm. They are able to build this device right into the iterations. This cannot be done in The Iterative Procedure. (It was tried but did not

work out well; it induced nonconvergence.) However, it is easy to see how to get around the problem. The new likelihood equations are just the equations for the following new model.

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \sum_{\substack{j=0 \\ j \notin S}}^{p_1} \underset{\sim}{U}_j \underset{\sim}{b}_j ,$$

where $\underset{\sim}{U}_0 = \underset{\sim}{I}_n$ and $\underset{\sim}{b}_0 = \underset{\sim}{e}$ and the other $\underset{\sim}{U}_j$ and $\underset{\sim}{b}_j$ are as defined in Section 1.3. But this model has exactly the same form as the old model for which The Iterative Procedure gave negative estimates; it is just smaller. Thus when The Iterative Procedure yields negative estimates, all that must be done is to reformulate the reduced problem above and resubmit this to The Iterative Procedure. One continues in this manner until no negative estimates are obtained. This procedure avoids entirely any negative estimates being reported as maximum likelihood estimates. The computer program presented in Appendix C does not automatically perform the above operations, but with a simple but major overhauling of the program, this could be accomplished. Then The Iterative Procedure would be comparable to the Hartley-Rao-Vaughn algorithm.

One serious problem of negative estimates in The Iterative Procedure was alluded to in Section 5.1. The problem is, that if at any stage, a negative estimate for any of the σ_i's occurs, then the matrix $\underset{\sim}{\Sigma}(\underset{\sim}{\sigma})$ may be singular or nearly singular. When this happens The Iterative Procedure will become very unstable and may even blow up. In

fact, this actually occurred in the Monte Carlo studies of The Iterative Procedure and was the only cause of nonconvergence of The Iterative Procedure (See Section 5.8.). Unfortunately no technique exists at the moment for eliminating these difficulties. The subject is undergoing further study.

Even with these difficulties The Iterative Procedure still performs very well in comparison with its competition, the Hartley-Rao-Vaughn algorithm. Some comparisons are given in the next section.

5.7. Comparison of The Iterative Procedure with the Hartley-Rao-Vaughn Algorithm

Hartley and Vaughn (1972:135-144) give four examples of maximum likelihood estimation involving real data. These examples will not be reproduced here; the performance of the Hartley-Rao-Vaughn (to be abbreviated H-R-V for the rest of this chapter) algorithm as stated in Hartley and Vaughn and The Iterative Procedure (to be abbreviated TIP) on these problems. These comparisons will point out the computational efficiency of The Iterative Procedure. The H-R-V algorithm computes variance ratios (that is, it computes σ_0 and $\gamma_i = \frac{\sigma_i}{\sigma_0}$ for $i=1,2,\ldots,p_1$)

instead of variances. The results of TIP have been converted to this form for comparison. The problems were fed into the computer program described in Appendix C using the same initial guesses Hartley and Vaughn used (converted of course to variances). Efficiency will be measured by comparing the number of inversions[1] of a matrix of the form $\underset{\sim}{\Sigma}(\underset{\sim}{\sigma})$ which are required (each method must perform such calculations). Both methods require many other calculations, with the H-R-V requiring many more, but these inversions are the major computational effort. The Iterative Procedure requires either one or two inversions per iteration (See Section 5.8 and Appendix C.) while the H-R-V algorithm requires $\frac{(p_1+1)(p_1+2)}{2} + 1$ inversions for each iteration to make its approximation (Hartley and Vaughn [1972:133]). It will be seen that each procedure requires approximately the same number of iterations to converge, with consequent great savings for The Iterative Procedure in computational effort. The results of the four sample problems given by Hartley and Vaughn were as follows.

1) The Twofold Nested Model

$$y_{ijk} = \mu + a_i + b_{ij} + e_{ijk}, \quad i=1,2,\ldots,I, \quad j=1,2,\ldots,J, \quad k=1,2,\ldots,K,$$

1. Of course the inversion may be only done implicitly as in solving a number of linear equations.

where μ is fixed and the a's, b's and e's are the random effects. Such a model is described in Section 6.3 with the a's as fixed effects. For this model, there exist exact solutions and so as in Section 5.5, The Iterative Procedure achieves them in one iteration. (In the sample problem of Hartley and Vaughn, I=4, J=3, K=2, for n=24.) However, the computer requires one more iteration to recognize convergence, so TIP requires two iterations to achieve the final result, which agrees with the exact solution. A total of two inversions of the matrix $\underset{\sim}{\Sigma}(\underset{\sim}{\sigma})$ are required. The H-R-V algorithm required three complete cycles and hence required a total of $3\cdot\left[\frac{(2+1)(2+2)}{2}+1\right] = 21$ inversions to obtain answers which agreed with the exact answers to only two decimal places for the variance ratios. Of course, had a more stringent convergence criterion been applied, H-R-V would have attained better agreement with the exact answers at a cost of more iterations.

The final results were

	$\hat{\sigma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
Exact	0.066542	39.106	24.204
H-R-V	0.066549	39.095	24.1999
TIP	0.066542	39.106	24.204

2) Twofold Nested Model When One Variance Ratio is Zero

This model is the same as above but for these data the variance ratio for the a effect is calculated as negative and hence set to zero.

(In the sample problem of Hartley and Vaughn, I=5, J=2, K=2, for n=20.) The Iterative Procedure requires two iterations to recognize the negative variance and two more to compute the final answer in the reduced model as described in Section 5.6. Thus a total of four inversions were required. The H-R-V algorithm required four complete cycles and hence 4·7=28 inversions to obtain the final answers. Both answers agreed with the exact answers.

The final results were

	$\hat{\sigma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
Exact	0.0387	0.0	0.35695
H-R-V	0.0387	0.0	0.35696
TIP	0.0387	0.0	0.35695 .

3) Unbalanced One-Way Classification

$$y_{ij} = \mu + a_i + e_{ij}, \quad i=1,2,\ldots,I, \quad j=1,2,\ldots,J_i.$$

This model is discussed in Section 6.4. The a's and e's are random effects and μ is a fixed effect. No exact solutions exist here. (In the sample problem of Hartley and Vaughn, I=5 and the J_i's are 5,3,2,3,1, for n=14.) The Iterative Procedure requires five iterations involving a total of 6 inversions to converge. The H-R-V algorithm required five iterations with a consequent $5\cdot\left(\frac{(1+1)(1+2)}{2} + 1\right) = 20$ inversions. The answers agreed to three decimal places. (Hartley and Vaughn only gave

the answers to three decimal places in this problem.)

The final results were

	$\hat{\sigma}_0$	$\hat{\gamma}_1$
H-R-V	0.773	0.671
TIP	0.773	0.671

4) Two-Way Classification with Interaction

$$y_{ijk} = \mu + a_i + b_j + c_{ij} + e_{ijk}\,, \quad \begin{array}{l} i=1,2,\ldots,I, \\ j=1,2,\ldots,J, \\ k=1,2,\ldots,K, \end{array}$$

where μ is a fixed effect and the a's, b's, c's and e's are random effects. This model is discussed in Sections 6.1 and 6.2. No closed form solutions exist for this model. (In the sample problem of Hartley and Vaughn I=2, J=3, K=3, for n=18.) The Iterative Procedure required 12 iterations involving a total of 13 inversions to reach a solution. The H-R-V algorithm required 15 iterations with a consequent

$15\cdot\left(\frac{(3+1)(3+2)}{2} + 1\right) = 165$ inversions. The answers for the two procedures only agree to two or three digits but as far as is possible to know The Iterative Procedure gives more accurate results. (For example $\hat{\sigma}_0$ can be solved for exactly in this model; see Section 6.2.) Again the H-R-V algorithm could be made to converge more closely, but at the cost of more iterations.

The final results were

	$\hat{\sigma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
H-R-V	69.75	9.54	12.82	0.326
TIP	69.78	9.52	12.79	0.327

These four examples point out that The Iterative Procedure seems to be computationally much more efficient than the H-R-V algorithm. To be sure, it may be unfair to compare in cases 1) and 2) where The Iterative Procedure gives exact solutions and the H-R-V just iterates. Still, it is an advantage to give exact solutions when they exist. In the cases where exact solutions do not exist the comparison is just as dramatic. One can hardly generalize from four examples, but it is true that a general pattern develops that if one method requires many iterations, so will the other; however, the H-R-V algorithm pays a much higher price per iteration. This pattern continues in the Monte Carlo studies in the next section. The H-R-V method has an advantage that it has been proven to converge. Such a result has not been proved for The Iterative Procedure but Monte Carlo results have been most encouraging. These are presented in the next section.

5.8. Monte Carlo Results

As noted in previous sections, it has not been proved here that The Iterative Procedure is guaranteed to converge. Thus an attempt has been made to demonstrate its effectiveness by Monte Carlo methods. The results have been most encouraging. The following procedure was used. Only one model, the two-way crossed balanced design described in Sections 6.1 and 6.2, was used. There were two reasons for this. First, the iterative equations can be easily programmed and do not require massive amounts of computer time to reach a solution as might be required with more complicated designs. Second, there is no closed form solution to these equations so the iterations are nontrivial. Problems which might possibly occur can occur under this setup as easily as under any other. This layout seemed to offer the best possibility for finding whatever problems The Iterative Process might have in the most economical way.

The actual Monte Carlo process was carried out as follows. Sufficient statistics exist for this model (see Section 6.1) and can be generated with pseudorandom chi-square generators once I, J, K and a set of "true" parameters σ_0, σ_1, σ_2, σ_3 have been chosen. These statistics were then fed into The Iterative Process. If convergence occurred, any negative estimates were set to zero and the equations resolved with those estimates restricted. When final solutions were obtained they were recorded and fed into the Hartley-Rao-Vaughn algorithm to confirm that they were indeed solutions of the likelihood equations. If 300 iterations occurred without convergence the process

stopped and the sufficient statistics were recorded and later fed into another program to analyze in more detail why convergence failed to occur. Over 15,000 separate Monte Carlo repetitions were run for various I, J, K, σ_0, σ_1, σ_2, σ_3 combinations. (See below for more detail on the types of combinations used.)

The results of the Monte Carlo analysis support the contention that The Iterative Procedure is indeed an effective, efficient method for calculating maximum likelihood estimates in the mixed model of the analysis of variance. Two problems were discovered during the Monte Carlo runs, one of which was easily rectified and the other of which was more serious. These problems will be discussed next followed by the large quantity of favorable results of the Monte Carlo trials.

When the Monte Carlo runs were started it was found that sometimes it seemed like The Iterative Process was not converging when in fact it was. (This occurred less than 0.8% of the time.) The problem was that the convergence was so slow that it might have taken up to 10,000 iterations to converge. For instance the sequence 11000, 9000, 10999, 9001, 10998, 9002,... is converging to 10000 but will require 2000 iterations to get there at the rate it is going. The Monte Carlo results seemed to indicate that whenever the slow convergence occurred, it was in the oscillating manner indicated above. A natural correction for this problem seemed to be to average consecutive iterations. This proved to be an excellent idea as it eliminated entirely the problem of reporting lack of convergence due to slow convergence and improved other slow convergence rates by a factor of from 10 to 50 or more. However,

care must be taken in applying this technique since if applied indiscriminantly, it can lead to a false convergence where a sequence of iterates regenerate themselves by averaging. (This actually occurred in early studies.) The false convergence can be easily eliminated by simple programming steps which are described in Appendix C. All further results reported here are results after the above modification was made to The Iterative Process.

A more serious problem which occurred was mentioned briefly in Section 5.6. When negative estimates were generated in one iteration in such a configuration that $\underset{\sim}{\Sigma}(\underset{\sim}{\sigma})$ became singular or nearly singular, the process became very unstable often oscillating between several points of singularity. Once such unstable points were reached convergence almost never occurred. No totally satisfactory solution to this problem was found. However, some facts became evident. First, the problem can only occur when some of the estimates are negative; for positive estimates convergence occurred 100% of the time. Second, the problem tended to occur only in situations when I, J and K were small and when σ_0, σ_1, σ_2, and σ_3 were chosen to increase the probability of negative estimates occurring. (For instance if the interaction variance σ_1 is large with respect to the row and column effect variances σ_2 and σ_3 or if the error variance σ_0 is large with respect to σ_1 then the probability of this problem occurring seemed to increase.) It is somewhat reassuring to note that such configurations occur infrequently in real data.

Although the above problem is serious, it may possibly be overcome by restarting from a new trial solution. If this fails it is possible to act as if the final solution had been reached at whatever point the blow up or oscillation occurred. Any estimates which are negative at this point are set to zero and the equations resolved. This second method is not totally satisfactory but will work in most cases.

The results of the Monte Carlo study not only pointed out two problems--one of which was completely rectifiable the other only partially so--but also produced many encouraging results. The first positive result was the fact that in over 15000 runs with various I, J, K, σ_0, σ_1, σ_2, σ_3 combinations, (see below for more detail) The Iterative Procedure converged 98.7% of the time when the failures due to slow convergence are included and 99+% of the time when the averaging correction is made. In every case where convergence occurred, the Hartley-Rao-Vaughn algorithm converged in one iteration starting with the final results of TIP as an initial guess. This verified that in every case where TIP converged, it converged to <u>a</u> solution of the likelihood equations.

Another positive result of the Monte Carlo study was that The Iterative Procedure seldom required a large number of iterations to reach convergence. The number of iterations seldom exceeded 20 although in a few cases it was over 200. The averaging modification mentioned above further reduced the number of iterations required. This contrasts with other procedures which may require large number of iterations. For instance, the Hartley-Rao-Vaughn algorithm may require a thousand or more

Runge-Kutta iterations for each approximate solution and there are as many of these approximate solutions as there are overall iterations. (It was for this reason that in the Monte Carlo study, the H-R-V algorithm was used only as a check rather than being allowed to iterate to a solution.)

It was also noted that if the "true" parameters were chosen so that negative estimates were unlikely to develop (see above), the percentage of runs where convergence occurred went up and the number of iterations went down. As long as no negative estimates occurred, The Iterative Procedure converged every time. It seems that such a configuration of parameters is likely to occur in real data.

Another very important aspect of The Iterative Procedure which was highlighted by the Monte Carlo study is that as the size of the design increased, so did the computational efficiency of TIP. The percentage of runs where convergence occurred increased (reaching 100% for large designs). The average number of iterations per run decreased (getting very close to the absolute minimum of 2 for large designs and attaining this minimum for very large designs).

The positive results mentioned above are illustrated by the data in the following two tables. In each table a box represents 100 Monte Carlo runs at the particular I, J, K and σ_0, σ_1, σ_2, σ_3 combination. The number in the upper left of each box is the percentage of runs on which convergence occurred and the number in the lower right is the average number of iterations per run for the runs represented by that box. Table 5.8.1 illustrates what occurs as n becomes large. Two

Table 5.8.1

Percentage of Runs on which Convergence Occurred and Average Number of Iterations Per Run on Sets of 100 Monte Carlo Runs with Various $I,J,K,\sigma_0,\sigma_1,\sigma_2,\sigma_3$ Combinations

I,J,K \ σ_0, σ_1, σ_2, σ_3	2 1 3 3	2 1 3 30	2 10 3 3	20 1 3 30	Average
(n=60) 5,4,3	99.0 4.77	100.0 3.98	98.0 7.56	99.0 4.84	99.0 5.28
(n=540) 15,12,3	100.0 2.53	100.0 2.47	100.0 2.85	100.0 2.59	100.0 2.71
Average	99.5 3.65	100.0 3.23	99.0 5.18	99.5 3.71	99.5 3.94

Table 5.8.2

Percentage of Runs on which Convergence Occurred and Average Number of Iterations Per Run on Sets of 100 Monte Carlo Runs with Various I,J,K,σ_0,σ_1,σ_2,σ_3 Combinations

I,J,K \ σ_0 / σ_1 / σ_2 / σ_3	1 / 1 / 3 / 2	10 / 3 / 30 / 2	10 / 1 / 30 / 25	10 / 10 / 10 / 10	10 / 7 / 40 / 20	Average
(n=18) 3,3,2	100.0 8.55	99.0 8.20	99.0 6.46	99.0 11.74	97.0 8.57	98.8 8.70
(n=27) 3,3,3	97.0 9.01	98.0 10.59	100.0 6.15	98.0 11.34	100.0 7.15	98.6 8.83
(n=24) 3,4,2	97.0 7.10	98.0 7.71	100.0 5.45	97.0 10.94	100.0 7.73	98.4 7.95
(n=36) 3,4,3	99.0 6.96	98.0 5.59	99.0 5.55	96.0 9.92	100.0 5.93	98.4 6.77
(n=32) 4,4,2	99.0 5.35	100.0 5.02	100.0 4.53	96.0 6.94	100.0 4.91	99.0 5.34
(n=40) 4,5,2	100.0 4.74	100.0 4.46	100.0 4.26	98.0 5.84	100.0 5.06	99.6 4.87
Average	98.7 6.94	98.8 6.92	99.7 5.40	97.3 9.46	99.5 6.55	98.8 7.04

different I, J, K combinations were studied--one with medium n=60 and one with large n=540. For large n the percentage of runs converging was 100% and the average number of iterations was only 2.61. (For the same σ_0, σ_1, σ_2, σ_3 combinations with n=6000, the average number of iterations was 2.0, the absolute minimum.) It is easily seen that some σ_0, σ_1, σ_2, σ_3 combinations are more difficult than others (in terms of number of iterations and to a smaller extent in percentage of runs converging) but that no combination is too difficult. This is probably accounted for by the fact that n is not small in either case. However, note that for all $\underset{\sim}{\sigma}$ combinations, there is a marked improvement upon going from the medium n to the large n case.

Table 5.8.2 illustrates results for several cases of small n ($\leq$ 40). It becomes apparent that there are no startling differences among the various I, J, K combinations although there seems to be a trend toward doing better if instead of n=IJK, only IJ is considered. (This leads to a suspicion that it may not be n but the m_i that are really important.) Increasing K did not seem to help. The trend is most easily seen in the decrease in the average number of iterations. Although there were not many differences in I, J, K combinations, there were noticeable differences among the σ_0, σ_1, σ_2, σ_3 combinations. Those for which σ_1 is not small relative to σ_2 and σ_3 (the first, second, and fourth) seem to do worse than the others both in terms of percentage of runs converging and average number of iterations. This agrees with the ideas advanced above. When the estimates were to be positive the convergence was swift and sure. One other pleasant aspect of Table 5.8.2 is that even for

these small designs the overall average percentage of runs converging was 98.8%.

As a summary of the Monte Carlo results the following can be said. It has not yet been proved in a theorem that any of the desirable properties mentioned in Section 5.4 are true for The Iterative Procedure. However, the following have been demonstrated.

1) The method usually converges (99+% of the time) and when it does it converges to a solution of the likelihood equations.
2) As the size of the design gets larger the method becomes more computationally efficient and even for small designs it is quite efficient.
3) If the parameters being estimated are configured well in the sense mentioned above, as they often are in real data, then the method is more effective.
4) The method does have a problem with negative estimates which can be partially overcome by the methods mentioned above.

In short, although it is not perfect, The Iterative Procedure seems to be a highly efficient, effective algorithm for solving the problems it was intended to solve.

CHAPTER 6

EXAMPLES

6.1. An Example of Application of Asymptotic Theory--The Two-Way Crossed Balanced Random Effects Model

The model used here is given first in the form conventionally used in the analysis of variance.

$$y_{ijk} = \mu + a_i + b_j + c_{ij} + e_{ijk},$$

$$i=1,2,\ldots,I,$$

$$j=1,2,\ldots,J,$$

$$k=1,2,\ldots,K.$$

Here μ is the overall mean effect and the a's, b's, c's and e's are random variables. Now list the y's in lexicographic order (see example below) as a vector to get the following model.

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \underset{\sim}{U}_1\underset{\sim}{b}_1 + \underset{\sim}{U}_2\underset{\sim}{b}_2 + \underset{\sim}{U}_3\underset{\sim}{b}_3 + \underset{\sim}{e},$$

where

$\underset{\sim}{y}$ is nX1, n = IJK,

$\underset{\sim}{X}$ is an nX1 vector of 1's,

$\underset{\sim}{\alpha}$ is a 1X1 unknown constant,

$\underset{\sim}{U}_1$ is an nXIJ standard design matrix for the AXB interactions,

$\underset{\sim}{U}_2$ is an nXI standard design matrix for the A effects,

$\underset{\sim}{U}_3$ is an nXJ standard design matrix for the B effects,

$\underset{\sim}{b}_1$ is an IJX1 random vector containing the AXB random interactions,

$\underset{\sim}{b}_2$ is an IX1 random vector containing the A random effects,

$\underset{\sim}{b}_3$ is a JX1 random vector containing the B random effects,

$\underset{\sim}{e}$ is an nX1 random vector containing the errors.

The parameters and constants in this model and their correspondence to the usual analysis of variance set up are given below. (ANOVA stands for analysis of variance.)

Parameter	Corresponding ANOVA Parameter
α	μ
σ_0	σ_e^2
σ_1	σ_{AB}^2
σ_2	σ_A^2
σ_3	σ_B^2

Constant	Corresponding ANOVA Constant
n	IJK
p_0	1
p_1	3
p	5
m_1	IJ
m_2	I
m_3	J

The actual form of the likelihood equations will be given later in this section. At this point $\underset{\sim}{y}$, $\underset{\sim}{X}$ $\underset{\sim}{U}_1$, $\underset{\sim}{U}_2$, $\underset{\sim}{U}_3$ for the case I=2, J=3, K=2 are given as illustrations.

$$\underset{\sim}{y}=\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} \quad \underset{\sim}{X}=\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \underset{\sim}{U}_1=\begin{bmatrix} 1&0&0&0&0&0 \\ 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&1&0 \\ 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \\ 0&0&0&0&0&1 \end{bmatrix} \quad \underset{\sim}{U}_2=\begin{bmatrix} 1&0 \\ 1&0 \\ 1&0 \\ 1&0 \\ 1&0 \\ 1&0 \\ 0&1 \\ 0&1 \\ 0&1 \\ 0&1 \\ 0&1 \\ 0&1 \end{bmatrix} \quad \underset{\sim}{U}_3=\begin{bmatrix} 1&0&0 \\ 1&0&0 \\ 0&1&0 \\ 0&1&0 \\ 0&0&1 \\ 0&0&1 \\ 1&0&0 \\ 1&0&0 \\ 0&1&0 \\ 0&1&0 \\ 0&0&1 \\ 0&0&1 \end{bmatrix}$$

The G_i matrices for i=0,2,3 are

$$\underset{\sim}{G}_0 = \begin{bmatrix} 1&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&1 \end{bmatrix} \quad \underset{\sim}{G}_1 = \begin{bmatrix} 1&1&0&0&0&0&0&0&0&0&0&0 \\ 1&1&0&0&0&0&0&0&0&0&0&0 \\ 0&0&1&1&0&0&0&0&0&0&0&0 \\ 0&0&1&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&1&0&0&0&0&0&0 \\ 0&0&0&0&1&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&1&0&0&0&0 \\ 0&0&0&0&0&0&1&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&1&1&0&0 \\ 0&0&0&0&0&0&0&0&1&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&1 \\ 0&0&0&0&0&0&0&0&0&0&1&1 \end{bmatrix}$$

$$\underset{\sim}{G}_2 = \begin{bmatrix} 1&1&1&1&1&1&0&0&0&0&0&0 \\ 1&1&1&1&1&1&0&0&0&0&0&0 \\ 1&1&1&1&1&1&0&0&0&0&0&0 \\ 1&1&1&1&1&1&0&0&0&0&0&0 \\ 1&1&1&1&1&1&0&0&0&0&0&0 \\ 1&1&1&1&1&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&1&1&1&1&1 \\ 0&0&0&0&0&0&1&1&1&1&1&1 \\ 0&0&0&0&0&0&1&1&1&1&1&1 \\ 0&0&0&0&0&0&1&1&1&1&1&1 \\ 0&0&0&0&0&0&1&1&1&1&1&1 \\ 0&0&0&0&0&0&1&1&1&1&1&1 \end{bmatrix} \qquad \underset{\sim}{G}_3 = \begin{bmatrix} 1&1&0&0&0&0&1&1&0&0&0&0 \\ 1&1&0&0&0&0&1&1&0&0&0&0 \\ 0&0&1&1&0&0&0&0&1&1&0&0 \\ 0&0&1&1&0&0&0&0&1&1&0&0 \\ 0&0&0&0&1&1&0&0&0&0&1&1 \\ 0&0&0&0&1&1&0&0&0&0&1&1 \\ 1&1&0&0&0&0&1&1&0&0&0&0 \\ 1&1&0&0&0&0&1&1&0&0&0&0 \\ 0&0&1&1&0&0&0&0&1&1&0&0 \\ 0&0&1&1&0&0&0&0&1&1&0&0 \\ 0&0&0&0&1&1&0&0&0&0&1&1 \\ 0&0&0&0&1&1&0&0&0&0&1&1 \end{bmatrix}$$

The asymptotic theory for this model will now be set up. Let I and J both approach infinity in such a way that $\frac{I}{J} \to \rho$, $0 < \rho < \infty$; K is fixed. Then the matrix of the ρ_{ij} defined in Section 4.2 is

$$\begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 1 & K & +\infty & +\infty \\ \frac{1}{K} & 1 & +\infty & +\infty \\ 0 & 0 & 1 & \rho \\ 0 & 0 & \frac{1}{\rho} & 1 \end{bmatrix} \end{array} .$$

It can be seen by inspection of the above matrix that c=2 and the sets S_s, s=0,1,2 are $S_0 = \{0\}$, $S_1\{1\}$, $S_2 = \{2,3\}$, where c and the S_s are defined in Section 4.2. Then it can be seen that

$$\text{rank}(\underset{\sim}{U}_1 : \underset{\sim}{U}_2 : \underset{\sim}{U}_3) = IJ \ ,$$

$$\text{rank}(\underset{\sim}{U}_2 : \underset{\sim}{U}_3) = I + J-1 \ ,$$

$$\text{rank}(\underset{\sim}{U}_2) = I \ ,$$

$$\text{rank}(\underset{\sim}{U}_3) = J \ .$$

Therefore

$$\begin{aligned} \nu_0 &= n - \text{rank}(\underset{\sim}{U}_1 : \underset{\sim}{U}_2 : \underset{\sim}{U}_3) \\ &= IJK - IJ \\ &= IJ(K-1), \end{aligned}$$

$$\begin{aligned} \nu_1 &= \text{rank}(\underset{\sim}{U}_1 : \underset{\sim}{U}_2 : \underset{\sim}{U}_3) - \text{rank}(\underset{\sim}{U}_2 : \underset{\sim}{U}_3) \\ &= IJ - (I + J - 1) \\ &= IJ - I - J + 1 \\ &= (I-1)(J-1), \end{aligned}$$

$$\begin{aligned} \nu_2 &= \text{rank}(\underset{\sim}{U}_2 : \underset{\sim}{U}_3) - \text{rank}(\underset{\sim}{U}_3) \\ &= I + J - 1 - J \\ &= I - 1, \end{aligned}$$

$$\begin{aligned} \nu_3 &= \text{rank}(\underset{\sim}{U}_2 : \underset{\sim}{U}_3) - \text{rank}(\underset{\sim}{U}_2) \\ &= I + J - 1 - I \\ &= J - 1. \end{aligned}$$

For α, $\nu_{p_1+1} = \nu_4$. It turns out that either I or J or any linear combination of the two will work. Arbitrarily, choose $\nu_4 = I$. Notice that $\frac{\nu_i}{m_i}$ approaches a reasonable limit in all cases here.

It can be shown with some work that Assumption 4.2.4 is true here. In fact tr $\underset{\sim}{D}_2^{-1}\underset{\sim}{U}_2'\underset{\sim}{U}_3\underset{\sim}{U}_3'\underset{\sim}{U}_2\underset{\sim}{D}_2^{-1} = \frac{I}{J} = (\frac{1}{J})\ I$ and tr $\underset{\sim}{D}_3^{-1}\underset{\sim}{U}_3'\underset{\sim}{U}_2\underset{\sim}{U}_2'\underset{\sim}{U}_3\underset{\sim}{D}_3^{-1} = \frac{J}{I} = (\frac{1}{I})\ J$. Thus in each case R_1 may be taken to be zero with R_2 equal to either $\frac{1}{J}$ or $\frac{1}{I}$ both of which converge to zero. This is because this experimental design is orthogonal in the sense of experimental design.

Now the matrices $\underset{\sim}{H}_s$ and $\underset{\sim}{P}$ defined in Section A.1 are defined for the example given earlier in this section. One of the many possible representations of $\underset{\sim}{P}$ is

$$\underset{\sim}{P} = \frac{1}{\sqrt{24}} \left[\underbrace{\begin{matrix} \sqrt{12} & 0 & 0 & 0 & 0 & 0 \\ -\sqrt{12} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{12} & 0 & 0 & 0 & 0 \\ 0 & -\sqrt{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{12} & 0 & 0 & 0 \\ 0 & 0 & -\sqrt{12} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{12} & 0 & 0 \\ 0 & 0 & 0 & -\sqrt{12} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{12} & 0 \\ 0 & 0 & 0 & 0 & -\sqrt{12} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{12} \\ 0 & 0 & 0 & 0 & 0 & -\sqrt{12} \end{matrix}}_{\underset{\sim}{H}_0} \quad \underbrace{\begin{matrix} 2 & -1 \\ 2 & -1 \\ -1 & 2 \\ -1 & 2 \\ -1 & -1 \\ -1 & -1 \\ -2 & 1 \\ -2 & 1 \\ 1 & -2 \\ 1 & -2 \\ 1 & 1 \\ 1 & 1 \end{matrix}}_{\underset{\sim}{H}_1} \quad \underbrace{\begin{matrix} \sqrt{2} & 2 & 0 & \sqrt{2} \\ \sqrt{2} & 2 & 0 & \sqrt{2} \\ \sqrt{2} & -1 & \sqrt{3} & \sqrt{2} \\ \sqrt{2} & -1 & \sqrt{3} & \sqrt{2} \\ \sqrt{2} & -1 & -\sqrt{3} & \sqrt{2} \\ \sqrt{2} & -1 & -\sqrt{3} & \sqrt{2} \\ -\sqrt{2} & 2 & 0 & \sqrt{2} \\ -\sqrt{2} & 2 & 0 & \sqrt{2} \\ -\sqrt{2} & -1 & \sqrt{3} & \sqrt{2} \\ -\sqrt{2} & -1 & \sqrt{3} & \sqrt{2} \\ -\sqrt{2} & -1 & -\sqrt{3} & \sqrt{2} \\ -\sqrt{2} & -1 & -\sqrt{3} & \sqrt{2} \end{matrix}}_{\underset{\sim}{H}_2} \right]$$

The columns making up $\underset{\sim}{H}_0$, $\underset{\sim}{H}_1$, and $\underset{\sim}{H}_2$ are noted under the matrix $\underset{\sim}{P}$. It

It is easily seen that

$$\text{rank}(\underset{\sim}{H}_0) = IJ(K-1) = 2\cdot 3\cdot 1 = 6,$$

$$\text{rank}(\underset{\sim}{H}_1) = (I-1)(J-1) = 1\cdot 2 = 2,$$

and

$$\text{rank}(\underset{\sim}{H}_2) = I+J-1 = 2+3-1 = 4.$$

It is also easy to verify that the following matrices are zero matrices.

$$\underset{\sim}{U}_1'\underset{\sim}{H}_0 = \underset{\sim}{0}, \quad \underset{\sim}{U}_2'\underset{\sim}{H}_0 = \underset{\sim}{0}, \quad \underset{\sim}{U}_3'\underset{\sim}{H}_0 = \underset{\sim}{0},$$

$$\underset{\sim}{G}_1\underset{\sim}{H}_0 = \underset{\sim}{0}, \quad \underset{\sim}{G}_2\underset{\sim}{H}_0 = \underset{\sim}{0}, \quad \underset{\sim}{G}_3\underset{\sim}{H}_0 = \underset{\sim}{0},$$

$$\underset{\sim}{H}_2'\underset{\sim}{H}_1 = \underset{\sim}{0}, \quad \underset{\sim}{U}_3'\underset{\sim}{H}_1 = \underset{\sim}{0}$$

$$\underset{\sim}{G}_2\underset{\sim}{H}_1 = \underset{\sim}{0}, \quad \underset{\sim}{G}_3\underset{\sim}{H}_1 = \underset{\sim}{0}.$$

The above illustrates the concepts in Section 4.2 for this case. Now proceed to actually write down the likelihood equations for this model. It turns out that all the $\underset{\sim}{G}_i$ matrices can be diagonalized by a single orthogonal matrix and that $\underset{\sim}{X}$ is a characteristic vector of each $\underset{\sim}{G}_i$. As noted in Section 5.3, this greatly simplifies the likelihood equations. In fact, $\underset{\sim}{G}_0 = (\underset{\sim}{I}_I \times \underset{\sim}{I}_J \times \underset{\sim}{I}_K)$, $\underset{\sim}{G}_1 = (\underset{\sim}{I}_I \times \underset{\sim}{I}_J \times \underset{\sim}{E}_K)$, $\underset{\sim}{G}_2 = (\underset{\sim}{I}_I \times \underset{\sim}{E}_J \times \underset{\sim}{E}_K)$, and $\underset{\sim}{G}_3 = (\underset{\sim}{E}_I \times \underset{\sim}{I}_J \times \underset{\sim}{E}_K)$, where $\times$ denotes the Kroenecker or direct product of two matrices and $\underset{\sim}{I}_I$, $\underset{\sim}{I}_J$ and $\underset{\sim}{I}_K$ are identity matrices and $\underset{\sim}{E}_I$, $\underset{\sim}{E}_J$ and $\underset{\sim}{E}_K$ are square matrices whose elements are all 1 of sizes I, J and K respectively. The matrix that diagonalizes

them all simultaneously is $\underset{\sim}{P}=(\underset{\sim}{P}_I \times \underset{\sim}{P}_J \times \underset{\sim}{P}_K)$ where $\underset{\sim}{P}_I$, $\underset{\sim}{P}_J$ and $\underset{\sim}{P}_K$ are any orthogonal matrices with first column proportional to a vector of ones of size I, J and K respectively.

Using the above diagonalization and the fact that $\underset{\sim}{X}$ is a characteristic vector of each $\underset{\sim}{G}_i$ it can be shown that the estimate of $\underset{\sim}{\alpha}$ can be made independent of the estimate of $\underset{\sim}{\sigma}$. The likelihood equations are written using the following terminology from the analysis of variance.

$$y_{...} = \frac{1}{IJK} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} y_{ijk} \qquad y_{i..} = \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} y_{ijk}$$

$$y_{.j.} = \frac{1}{IK} \sum_{j=1}^{J} \sum_{k=1}^{K} y_{ijk} \qquad y_{ij.} = \frac{1}{K} \sum_{k=1}^{K} y_{ijk}$$

$$SS_T = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{ijk} - y_{...})^2 \qquad SS_e = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{ijk} - y_{ij.})^2$$

$$SS_A = JK \sum_{i=1}^{I} (y_{i..} - y_{...})^2 \qquad SS_B = IK \sum_{j=1}^{J} (y_{.j.} - y_{...})^2$$

$$SS_{AB} = K \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij.} - y_{i..} - y_{.j.} + y_{...})^2$$

The likelihood equations are

$$\frac{IJK}{(\sigma_0 + K\sigma_1 + JK\sigma_2 + IK\sigma_3)} \alpha = \frac{IJK}{(\sigma_0 + K\sigma_1 + JK\sigma_2 + IK\sigma_3)} y_{...} ,$$

$$\frac{1}{\sigma_0+K\sigma_1+JK\sigma_2+IK\sigma_3}+\frac{I-1}{\sigma_0+K\sigma_1+JK\sigma_2}+\frac{J-1}{\sigma_0+K\sigma_1+IK\sigma_3}+\frac{(I-1)(J-1)}{\sigma_0+K\sigma_1}+\frac{IJ(K-1)}{\sigma_0}$$

$$=\frac{SS_A}{(\sigma_0+K\sigma_1+JK\sigma_2)^2}+\frac{SS_B}{(\sigma_0+K\sigma_1+IK\sigma_3)^2}+\frac{SS_{AB}}{(\sigma_0+K\sigma_1)^2}+\frac{SS_e}{\sigma_0^2}\ ,$$

$$\frac{K}{\sigma_0+K\sigma_1+JK\sigma_2+IK\sigma_3}+\frac{K(J-1)}{\sigma_0+K\sigma_1+JK\sigma_2}+\frac{K(I-1)}{\sigma_0+K\sigma_1+JK\sigma_3}+\frac{K(I-1)(J-1)}{\sigma_0+K\sigma_1}$$

$$=\frac{K(SS_A)}{(\sigma_0+K\sigma_1+JK\sigma_2)^2}+\frac{K(SS_B)}{(\sigma_0+K\sigma_1+IK\sigma_3)^2}+\frac{K(SS_{AB})}{(\sigma_0+K\sigma_2)^2}\ ,$$

$$\frac{JK}{\sigma_0+K\sigma_1+JK\sigma_2+IK\sigma_3}+\frac{JK(I-1)}{\sigma_0+K\sigma_1+JK\sigma_2}=\frac{JK(SS_A)}{(\sigma_0+K\sigma_1+JK\sigma_2)^2}\ ,$$

$$\frac{IK}{\sigma_0+K\sigma_1+JK\sigma_2+IK\sigma_3}+\frac{IK(J-1)}{\sigma_0+K\sigma_1+IK\sigma_3}=\frac{IK(SS_B)}{(\sigma_0+K\sigma_1+IK\sigma_3)^2}\ .$$

No closed form solution exists. The iterative solution is given in Section 6.2. The matrix $\underset{\sim}{J}$ defined by

$(\underset{\sim}{J})_{ij} = \lim \frac{1}{n_i n_j}\left[\mathcal{E}_0\left(-\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\theta})}{\partial\theta_i\partial\theta_j}\bigg|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0}\right)\right]$ is calculated in the following manner.

It is true that

$$\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X} = \frac{IJK}{\sigma_{00}+K_{01}+JK\sigma_{02}+IK\sigma_{03}} \; .$$

Therefore

$$\frac{1}{n_4^2}\,\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X} = \frac{1}{I} \cdot \frac{IJK}{\sigma_{00}+K\sigma_{01}+JK_{02}+IK\sigma_{03}}$$

$$\rightarrow \frac{1}{\sigma_{02} + \rho\sigma_{03}} \; .$$

Now $\underset{\sim}{C}_1$ whose $(i,j)^{th}$ element is given by $\frac{1}{2} \lim \frac{1}{n_i n_j} \operatorname{tr} \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j$,

$i,j=0,1,2,3$, must be calculated. The following table will suffice.

Table 6.1.1

Definition and Limits for the Elements of $\underset{\sim}{C}_1$

i	j	$\frac{1}{n_i n_j} \operatorname{tr} \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j$	Limit of $\frac{1}{n_i n_j} \operatorname{tr} \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j$
0	0	$\frac{1}{IJ(K-1)}\left[\frac{1}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02}+IK\sigma_{03})^2} + \frac{(I-1)}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02})^2} + \frac{(J-1)}{(\sigma_{00}+K\sigma_{01}+IK\sigma_{03})^2} + \frac{(I-1)(J-1)}{(\sigma_{00}+K\sigma_{01})^2} + \frac{IJ(K-1)}{\sigma_{00}^2}\right]$	$\frac{1}{\sigma_{00}^2} + \frac{1}{(K-1)(\sigma_{00}+K_{01})^2}$

i	j	$\frac{1}{n_i n_j}$ tr $\Sigma_0^{-1} G_i \Sigma_0^{-1} G_j$	Limit of $\frac{1}{n_i n_j}$ tr $\Sigma_0^{-1} G_i \Sigma_0^{-1} G_j$
0	1	$\frac{1}{[IJ(K-1)(I-1)(J-1)]^{\frac{1}{2}}}\left[\frac{K}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02}+IK\sigma_{03})^2} + \frac{K(I-1)}{(\sigma_{00}+K_{01}+JK\sigma_{02})^2} + \frac{K(J-1)}{(\sigma_{00}+K\sigma_{01}+IK\sigma_{03})^2} + \frac{K(I-1)(J-1)}{(\sigma_{00}+K\sigma_{01})^2}\right]$	$\frac{K}{(K-1)^{\frac{1}{2}}(\sigma_{00}+K\sigma_{01})^2}$
0	2	$\frac{1}{[IJ(K-1)(I-1)]^{\frac{1}{2}}}\left[\frac{JK}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02}+IK\sigma_{03})^2} + \frac{JK(I-1)}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02})^2}\right]$	0
0	3	$\frac{1}{[IJ(K-1)(J-1)]^{\frac{1}{2}}}\left[\frac{IK}{(\sigma_{00}+K_{01}+JK\sigma_{02}+IK\sigma_{03})^2} + \frac{IK(J-1)}{(\sigma_{00}+K\sigma_{01}+IK\sigma_{03})^2}\right]$	0
1	1	$\frac{1}{(I-1)(J-1)}\left[\frac{K^2}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02}+IK_{03})^2} + \frac{K^2(I-1)}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02})^2} + \frac{K^2(J-1)}{(\sigma_{00}+K\sigma_{01}+IK\sigma_{03})^2}\right.$	$\frac{K^2}{(\sigma_{00}+K\sigma_{01})^2}$

i	j	$\frac{1}{n_i n_j}\,\mathrm{tr}\,\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j$	Limit of $\frac{1}{n_i n_j}\,\mathrm{tr}\,\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j$
		$+\frac{K^2(I-1)(J-1)}{(\sigma_{00}+K\sigma_{01})^2}\Big]$	
1	2	$\frac{1}{[(I-1)^2(J-1)]^{\frac{1}{2}}}\Big[\frac{JK^2}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02}+IK\sigma_{03})^2}$ $+\frac{JK^2(I-1)}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02})^2}\Big]$	0
1	3	$\frac{1}{[(I-1)(J-1)^2]^{\frac{1}{2}}}\Big[\frac{IK^2}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02}+IK\sigma_{03})^2}$ $+\frac{IK^2(J-1)}{(\sigma_{00}+K\sigma_{01}+IK\sigma_{03})^2}\Big]$	0
2	2	$\frac{1}{(I-1)}\Big[\frac{J^2K^2}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02}+IK\sigma_{03})^2}$ $+\frac{J^2K^2(I-1)}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02})^2}\Big]$	$\frac{1}{\sigma_{02}^2}$
2	3	$\frac{1}{[(I-1)(J-1)]^{\frac{1}{2}}}\Big[\frac{IJK^2}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02}+IK\sigma_{03})^2}\Big]$	0
3	3	$\frac{1}{(J-1)}\Big[\frac{I^2K^2}{(\sigma_{00}+K\sigma_{01}+JK\sigma_{02}+IK\sigma_{03})^2}$	$\frac{1}{\sigma_{03}^2}$

i	j	$\frac{1}{n_i n_j} \operatorname{tr} \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j$	Limit of $\frac{1}{n_i n_j} \operatorname{tr} \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j$

$$+ \frac{I^2K^2(J-1)}{(\sigma_{00}+K\sigma_{01}+IK\sigma_{03})^2} \Big]$$

Thus it follows that

$$\underset{\sim}{C}_1 = \begin{bmatrix} \frac{1}{2}\left[\frac{1}{\sigma_{00}^2} + \frac{1}{(K-1)(\sigma_{00}+K\sigma_{01})^2}\right] & \frac{1}{2}\left[\frac{K}{(K-1)^{\frac{1}{2}}(\sigma_{00}+K\sigma_{01})^2}\right] & 0 & 0 \\ \frac{1}{2}\left[\frac{K}{(K-1)^{\frac{1}{2}}(\sigma_{00}+K\sigma_{01})^2}\right] & \frac{1}{2}\left[\frac{K^2}{(\sigma_{00}+K\sigma_{01})^2}\right] & 0 & 0 \\ 0 & 0 & \frac{1}{2}\frac{1}{\sigma_{02}^2} & 0 \\ 0 & 0 & 0 & \frac{1}{2}\frac{}{\sigma_{03}^2} \end{bmatrix}.$$

Hence

$$\underset{\sim}{C}_1^{-1} = \begin{bmatrix} 2\sigma_{00}^2 & -\dfrac{2\sigma_{00}^2}{(K-1)^{\frac{1}{2}}K} & 0 & 0 \\[2em] -\dfrac{2\sigma_{00}^2}{(K-1)^{\frac{1}{2}}K} & 2\sigma_{01}^2 + 2\,\dfrac{\sigma_{00}^2}{K(K-1)} + 4\,\dfrac{\sigma_{00}\sigma_{01}}{K} & 0 & 0 \\[2em] 0 & 0 & 2\sigma_{02}^2 & 0 \\[2em] 0 & 0 & 0 & 2\sigma_{03}^2 \end{bmatrix} .$$

The $\underset{\sim}{C}_1^{-1}$ obtained above may be compared with the asymptotic covariance matrix obtained when the usual analysis of variance estimators are normalized by the same normalizing sequences. Recall that the usual analysis of variance estimators are

$$\tilde{\sigma}_0 = \frac{SS_e}{IJ(K-1)} \;, \qquad \tilde{\sigma}_2 = \frac{1}{JK}\left[\frac{SS_A}{(I-1)} - \frac{SS_{AB}}{(I-1)(J-1)}\right],$$

$$\tilde{\sigma}_1 = \frac{1}{K}\left[\frac{SS_{AB}}{(I-1)(J-1)} - \frac{SS_e}{IJ(K-1)}\right], \qquad \tilde{\sigma}_3 = \frac{1}{IK}\left[\frac{SS_B}{(J-1)} - \frac{SS_{AB}}{(I-1)(J-1)}\right],$$

and that

$$SS_e \sim \sigma_{00}\ \chi^2_{IJ(K-1)} \;, \qquad SS_A \sim (\sigma_{00}+K\sigma_{01}+JK\sigma_{02})\ \chi^2_{(I-1)} \;,$$

$$SS_{AB} \sim (\sigma_{00}+K\sigma_{01})\ \chi^2_{(I-1)(J-1)}, \qquad SS_B \sim (\sigma_{00}+K_{01}+IK\sigma_{03})\ \chi^2_{(J-1)} \;,$$

and that all sums of squares are independent. One discovers using the above that the asymptotic covariance matrix of the $\tilde{\sigma}$'s is just $\underset{\sim}{C}_1^{-1}$. Thus the maximum likelihood and the usual ANOVA estimates are asymptotically equivalent in this case. Note that in either the usual ANOVA situation or the maximum likelihood situation had an attempt been made to normalize each estimate by the same sequence (for instance $n^{\frac{1}{2}}$), something would have gone wrong no matter what sequence was tried. That is, one of the asymptotic variances would have been zero or infinity.

To make this example totally complete would require a demonstration that the various lemmae in Chapter 4 are true in this case (which can be done in a straightforward manner by brute force). This will not be done here. It has been shown how the concepts of Chapter 4 apply in this case. In the next section the actual computation of the estimates is illustrated for this model.

6.2. Example of the Iterative Procedure--The 2-Way Crossed Balanced Random Effects Model

The specifications for the model and the likelihood equations have already been developed in Section 6.1. The equation for $\hat{\alpha}$ may be solved independent of the estimates of the σ_i's yielding $\hat{\alpha} = y_{...}$. The equations needed to iterate and solve for $\hat{\sigma}_0$, $\hat{\sigma}_1$, $\hat{\sigma}_2$, $\hat{\sigma}_3$ are of the form $\underset{\sim}{B}(\underset{\sim}{\sigma}_{(k)})\underset{\sim}{\sigma}_{(k+1)} = \underset{\sim}{c}(\sigma_{(k)})$ where $\underset{\sim}{B}(\underset{\sim}{\sigma}_{(k)})$ is a 4x4 matrix whose elements are given in Table 6.1.1 with $\sigma_i^{(k)}$ substituted for σ_{0i}, i=0,1,2,3 and $\underset{\sim}{c}(\underset{\sim}{\sigma}_{(k)})$ is a 4x1 vector consisting of the right hand sides of the likelihood equations given in Section 6.1 with $\sigma_i^{(k)}$ substituted for σ_i, i=0,1,2,3. After much algebraic manipulation these equations can be solved to yield the following iterative equations. (Define

$$MS_e = \frac{SS_e}{IJ(K-1)}\ , \ MS_A = \frac{SS_A}{(I-1)}\ , \ MS_B = \frac{SS_B}{(J-1)}\ , \ MS_{AB} = \frac{SS_{AB}}{(I-1)(J-1)}\ .)$$

$$\sigma_0^{(k+1)} = MS_e$$

$$\sigma_1^{(k+1)} = \frac{1}{K}\left[MS_{AB}-MS_e\right] + \frac{X}{K}\left[(\sigma_0^{(k)}+K\sigma_1^{(k)})^2\right] \cdot \left[MS_A + MS_B - MS_{AB}\right]$$

$$\sigma_2^{(k+1)} = \frac{1}{JK}\left[MS_A-MS_{AB}\right] - \frac{X}{JK}\left[(\sigma_0^{(k)}+K\sigma_1^{(k)})^2+(J-1)(\sigma_0^{(k)}+K\sigma_1^{(k)}+JK\sigma_2^{(k)})^2\right] \cdot \left[MS_A + MS_B - MS_{AB}\right]$$

$$\sigma_3^{(k+1)} = \frac{1}{IK}\left[MS_B - MS_{AB}\right] - \frac{X}{IK}\left[(\sigma_0^{(k)} + K\sigma_1^{(k)})^2 + (I-1)(\sigma_0^{(k)} + K\sigma_1^{(k)} + IK\sigma_3^{(k)})^2\right]$$

$$\cdot \left[MS_A + MS_B - MS_{AB}\right]$$

where $X = \Big[(\sigma_0^{(k)} + K\sigma_1^{(k)})^2 + (J-1)(\sigma_0^{(k)} + K\sigma_1^{(k)} + JK\sigma_2^{(k)})^2$

$$+ (I-1)(\sigma_0^{(k)} + K\sigma_1^{(k)} + IK\sigma_3^{(k)})^2$$

$$+ (I-1)(J-1)(\sigma_0^{(k)} + K\sigma_1^{(k)} + JK\sigma_2^{(k)} + IK\sigma_3^{(k)})^2\Big]^{-1} .$$

Observe that at each stage the new iterated estimator is the usual analysis of variance-method of moments estimator plus a correction term. The correction term is a product of the quantity $MS_A+MS_B-MS_{AB}$ and a term which depends on the old estimates which were iterated. If in fact $MS_A+MS_B-MS_{AB}$ is zero then there is no correction and the maximum likelihood and usual estimates coincide. The reason for this is that if $MS_A+MS_B-MS_{AB}= 0$, the likelihood equations admit an explicit solution and therefore, as was shown in Section 5.5, the iterative process must yield this exact solution in one iteration. This procedure is easy to program on a computer or programmable calculator and may be used to easily compare the two types of estimates.

No closed form expression is possible for the maximum likelihood estimators so their behavior cannot be studied directly in this case. A case where explicit solutions do exist and where the behavior of the estimates can be studied directly is presented in the next section.

6.3. Example where Explicit Closed Form Solutions to the Likelihood Equations Exist--The Two-Way Balanced Nested Layout as a Mixed Model

This model, given in the usual analysis of variance notation, is

$$y_{ijk} = \alpha_i + b_{ij} + e_{ijk}$$

$$i=1,2,\ldots,I; \quad j=1,2,\ldots,J; \quad k=1,2,\ldots,K;$$

where y_{ijk} is the observation, α_i is an unknown fixed effect and the b_{ij} are independent, identically distributed as $\mathcal{N}(0,\sigma_B^2)$ and the e_{ijk} are independent, identically distributed as $\mathcal{N}(0,\sigma_e^2)$ and the b_{ij} and e_{ijk} are independent. Listing the y's in lexicographic order as in Section 6.1 the following model is obtained.

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \underset{\sim}{U}_1\underset{\sim}{b}_1 + \underset{\sim}{e},$$

where

$\underset{\sim}{y}$ is nX1, n=IJK,

$\underset{\sim}{X}$ is an nXI standard design matrix for A effects,

$\underset{\sim}{\alpha}$ is an IX1 vector of unknown constants,

$\underset{\sim}{U}_1$ is an nXIJ standard design matrix for interactions,

$\underset{\sim}{b}_1$ is an IJX1 random vector containing the random effects for this model,

$\underset{\sim}{e}$ is an nX1 random vector containing the errors.

The parameters and constants in this model and their corresponding parameters and constants in the usual analysis of variance set up are

Parameter	Corresponding ANOVA Parameter
α_i	α_i
σ_0	σ_e^2
σ_1	σ_B^2

Constant	Corresponding ANOVA Constant
n	IJK
p_0	I
p_1	1
p	I+2
m_1	IJ

$\underset{\sim}{y}$, $\underset{\sim}{X}$, $\underset{\sim}{U}_1$ are illustrated below for the case I=2, J=3, K=2.

$$\underset{\sim}{y} = \begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} \qquad \underset{\sim}{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \qquad \underset{\sim}{U}_1 = \begin{bmatrix} 1&0&0&0&0&0 \\ 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&1&0 \\ 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \\ 0&0&0&0&0&1 \end{bmatrix}$$

It can be shown that $\underset{\sim}{G}_0$ and $\underset{\sim}{G}_1$ are diagonalizable and that $\underset{\sim}{X}$ consists of characteristic vectors of $\underset{\sim}{G}_0$ and $\underset{\sim}{G}_1$ with each column of $\underset{\sim}{X}$ having the same characteristic value. Using this information it is easy to show that the likelihood equations are as follows. For $\underset{\sim}{\alpha}$, $(\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{X})\underset{\sim}{\alpha} = \underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{y}$ reduces to $(\sigma_0+K\sigma_1)^{-1}JK\alpha_i = (\sigma_0+K\sigma_1)^{-1}JKy_{i..}$, $i=1,2,\ldots,I$ ($y_{i..}$ defined in Section 6.1) which yields $\hat{\alpha}_i = y_{i..}$ independent of the choice of σ_0 or σ_1, $i=1,2,\ldots,I$. Let

$SS_B = K \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij.} - y_{i..})^2$. Then the equations for σ_0 and σ_1 are

$$\frac{IJ}{(\sigma_0+K\sigma_1)} + \frac{IJ(K-1)}{\sigma_0} = \frac{SS_B}{(\sigma_0+K\sigma_1)^2} + \frac{SS_e}{\sigma_0^2} \quad ,$$

$$\frac{IJK}{(\sigma_0+K\sigma_1)} = \frac{K(SS_B)}{(\sigma_0+K\sigma_1)^2} \quad .$$

It is easy to solve these equations to obtain

$$\hat{\sigma}_0 = \frac{SS_e}{IJ(K-1)} = MS_e \quad ,$$

$$\hat{\sigma}_1 = \frac{1}{K}\left[\frac{SS_B}{IJ} - MS_e\right] .$$

To show that the iteration $\underset{\sim}{B}(\underset{\sim}{\sigma}_{(k)})\underset{\sim}{\sigma}_{(k+1)} = \underset{\sim}{c}(\underset{\sim}{\sigma}_{(k)})$ yields these solutions in one iteration, set up $\underset{\sim}{B}(\underset{\sim}{\sigma}_{(k)})$ and $\underset{\sim}{c}(\underset{\sim}{\sigma}_{(k)})$ as follows.

$$\underset{\sim}{B}(\underset{\sim}{\sigma}_{(k)}) = \begin{bmatrix} \dfrac{IJ}{(\sigma_0^{(k)}+K\sigma_1^{(k)})^2} + \dfrac{IJ(K-1)}{(\sigma_0^{(k)})^2} & \dfrac{IJK}{(\sigma_0^{(k)}+K\sigma_1^{(k)})^2} \\ & \\ \dfrac{IJK}{(\sigma_0^{(k)}+K\sigma_1^{(k)})^2} & \dfrac{IJK^2}{(\sigma_0^{(k)}+K\sigma_1^{(k)})^2} \end{bmatrix}$$

$$\underset{\sim}{g}(\underset{\sim}{\sigma}_{(k)}) \doteq \begin{bmatrix} \dfrac{SS_B}{(\sigma_0^{(k)}+K\sigma_1^{(k)})^2} + \dfrac{SS_e}{(\sigma_0^{(k)})^2} \\ \\ \dfrac{K(SS_B)}{(\sigma_0^{(k)}+K\sigma_1^{(k)})^2} \end{bmatrix}$$

Solution of these equations for any $\sigma_0^{(k)}$ and $\sigma_1^{(k)}$ where $\sigma_0^{(k)} > 0$ and $\sigma_0^{(k)}+K\sigma_1^{(k)} \neq 0$ yields exactly $\hat{\sigma}_0$ and $\hat{\sigma}_1$ given above.

The asymptotic theory for this case is easy since the non-asymptotic distributions of all the estimators are known. It is well known that $SS_e \sim \sigma_{oo}\ \chi^2_{IJ(K-1)}$, $SS_B \sim (\sigma_{oo} + K\sigma_{o1})\ \chi^2_{I(J-1)}$, $\hat{\underset{\sim}{\alpha}} \sim \mathcal{N}_I(\underset{\sim}{\alpha}_o, \frac{(\sigma_{oo}+K\sigma_{o1})}{JK}\underset{\sim}{I})$, and that $\hat{\underset{\sim}{\alpha}}$ is independent of both SS_B and SS_e. From this it follows that $\hat{\underset{\sim}{\alpha}}$ and $\hat{\sigma}_0$ are unbiased but that $\hat{\sigma}_1$ is biased with expected value $\mathcal{E}_0(\hat{\sigma}_1) = (1 + \frac{1}{IJ})\sigma_{o1} - \frac{1}{IJ}\sigma_{oo}$. Furthermore, the variance-covariance matrix of $(\hat{\alpha}', \hat{\sigma}_0, \hat{\sigma}_1)'$ is

$$\begin{bmatrix} \dfrac{(\sigma_{oo}+K\sigma_{o1})}{JK}\underset{\sim}{I}_I & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0}' & \dfrac{2\sigma_{oo}^2}{IJ(K-1)} & -\dfrac{2\sigma_{oo}^2}{IJK(K-1)} \\ \underset{\sim}{0}' & -\dfrac{2\sigma_{oo}^2}{IJK(K-1)} & \dfrac{2}{IJK^2}\left[\dfrac{(J-1)(\sigma_{oo}+K\sigma_{o1})^2}{J} + \dfrac{\sigma_{oo}^2}{(K-1)}\right] \end{bmatrix}$$

Two cases of asymptotic behavior may now be considered. In both cases I must remain fixed because it is the number of fixed parameters. Since this is true, J must become infinite; otherwise $m_1 = IJ$ will not become infinite. The two cases then will be K fixed and K becomes infinite. To be sure, nothing is gained if K becomes infinite but it does introduce items of pedantic interest.

Case I: K fixed.

It is obvious either by examination of the variance-covariance matrix or by a return to definitions that the correct normalizing sequences are

$$n_0 = [IJ(K-1)]^{\frac{1}{2}},$$

$$n_1 = [IJ]^{\frac{1}{2}},$$

$$n_2 = [JK]^{\frac{1}{2}}.$$

Then $\underset{\sim}{J}^{-1}$, the limiting covariance matrix, becomes

$$\underset{\sim}{J}^{-1} = \begin{bmatrix} (\sigma_{00}+K\sigma_{01})\underset{\sim}{I}_I & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0}' & 2\sigma_{00}^2 & \dfrac{2\sigma_{00}^2}{K(K-1)^{\frac{1}{2}}} \\ \underset{\sim}{0}' & \dfrac{2\sigma_{00}^2}{K(K-1)^{\frac{1}{2}}} & \dfrac{2}{K^2}\left[(\sigma_{00}+K\sigma_{01})^2 + \dfrac{\sigma_{00}^2}{(K-1)}\right] \end{bmatrix}.$$

This matrix can be obtained either directly from the finite covariance matrix or by the definitions of Section 4.3. In either case the above matrix is obtained. Note that n_0 and n_1 are of the same order of

magnitude in this case. However in the second case this will not be so.

Case II: K becomes infinite.

In this case the correct normalizing sequences are

$$n_0 = [IJ(K-1)]^{\frac{1}{2}}$$

$$n_1 = [IJ]^{\frac{1}{2}}$$

$$n_2 = J^{\frac{1}{2}}$$

The limiting covariance matrix is then

$$\underset{\sim}{J}^{-1} = \begin{bmatrix} \sigma_{01}\underset{\sim}{I}_I & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0}' & 2\sigma_{00}^2 & 0 \\ \underset{\sim}{0}' & 0 & 2\sigma_{01}^2 \end{bmatrix}$$

In this case n_0 and n_1 are not of the same order of magnitude and so the asymptotic covariance between $\hat{\sigma}_0$ and $\hat{\sigma}_1$ becomes zero. In case I, normalizing all estimators by $n^{\frac{1}{2}} = [IJK]^{\frac{1}{2}}$ could work but in case II it would not work at all.

In both cases the maximum likelihood estimates are asymptotically equivalent to the usual analysis of variance-method of moments estimates as can easily be verified.

6.4. Example of an Unbalanced Layout--The One-Way Unbalanced Random Effects Model

The model, given in the usual analysis of variance notation, is

$$y_{ij} = \mu + a_i + e_{ij}, \quad j=1,2,\ldots,J_i, \quad i=1,2,\ldots,I;$$

where y_{ij} is the observation, μ is an unknown mean, the a_i are independent, identically distributed as $\mathcal{N}(0,\sigma_A^2)$ and the e_{ij} are independent, identically distributed as $\mathcal{N}(0,\sigma_e^2)$ and the a_i and e_{ij} are independent. Listing the y's in lexicographic order the following model is obtained.

$$\underset{\sim}{y} = \underset{\sim}{X}\alpha + \underset{\sim}{U}_1\underset{\sim}{b}_1 + \underset{\sim}{e},$$

where

$\underset{\sim}{y}$ is nX1, $\left(n = \sum_{i=1}^{I} J_i\right)$,

$\underset{\sim}{X}$ is an nX1 vector of ones,

α is a 1X1 unknown constant,

$\underset{\sim}{U}_1$ is an nXI standard design matrix for this model,

$\underset{\sim}{b}_1$ is an IX1 vector containing the random effects,

$\underset{\sim}{e}$ is an nX1 vector containing the errors.

The parameters and constants in this model and their corresponding analysis of variance parameters and constants are

Parameter	Corresponding ANOVA Parameter
α	μ
σ_0	σ_e^2
σ_1	σ_A^2

Constant	Corresponding ANOVA Constant
n	$\sum_{i=1}^{I} J_i$
p_0	1
p_1	1
p	3
m_1	I

The case $I=2$, $J_1=2$, $J_2=3$ is illustrated.

$$\underset{\sim}{y} = \begin{bmatrix} y_{11} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix}, \qquad \underset{\sim}{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \underset{\sim}{U}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

Let $y_{i.} = \dfrac{\sum_{j=1}^{J_i} y_{ij}}{J_i}$. The likelihood equations are as follows. For α the equation is

$$\left[\sum_{i=1}^{I} \frac{J_i}{(\sigma_0+J_i\sigma_1)}\right] \alpha = \sum_{i=1}^{I} \left[\frac{J_i y_{i.}}{(\sigma_0+J_i\sigma_1)^2}\right].$$

For σ_0 the equation is

$$\sum_{i=1}^{I} \left[\frac{1}{(\sigma_0+J_i\sigma_1)} + \frac{(J_i-1)}{\sigma_0}\right] = \sum_{i=1}^{I} \left[\frac{J_i(y_{i.}-\alpha)^2}{(\sigma_0+J_i\sigma_1)^2} + \frac{\sum_{j=1}^{J_i}(y_{ij}-y_{i.})^2}{\sigma_0^2}\right].$$

For σ_1 the equation is

$$\sum_{i=1}^{I} \left[\frac{J_i}{(\sigma_0+J_i\sigma_1)}\right] = \sum_{i=1}^{I} \left[\frac{J_i^2(y_{i.}-\alpha)^2}{(\sigma_0+J_i\sigma_1)^2}\right].$$

These equations are very messy. They must be solved simultaneously for α, σ_0 and σ_1 and no real simplifications are possible. The iteration equations are given by

$$\underset{\sim}{\sigma}_{(k+1)} = \underset{\sim}{B}^{-1}(\underset{\sim}{\sigma}_{(k)})\ \underset{\sim}{c}[\underset{\sim}{\sigma}_{(k)}, \underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(k)})], \quad k=1,2,\ldots ,$$

where $\underset{\sim}{B}(\underset{\sim}{\sigma}_{(k)})$ is the following 2X2 matrix

$$\underset{\sim}{B}(\underset{\sim}{\sigma}_{(k)})=\begin{bmatrix} \sum_{j=1}^{I}\left\{\frac{1}{[\sigma_0^{(k)}+J_j\sigma_1^{(k)}]^2}+\frac{(J_j-1)}{[\sigma_0^{(k)}]^2}\right\} & \sum_{j=1}^{I}\left\{\frac{J_j}{[\sigma_0^{(k)}+J_j\sigma_1^{(k)}]^2}\right\} \\ \sum_{j=1}^{I}\left\{\frac{J_j}{[\sigma_0^{(k)}+J_j\sigma_1^{(k)}]^2}\right\} & \sum_{j=1}^{I}\left\{\frac{J_j^2}{[\sigma_0^{(k)}+J_j\sigma_1^{(k)}]^2}\right\} \end{bmatrix},$$

and $\underset{\sim}{c}[\underset{\sim}{\sigma}_{(k)}, \underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(k)})]$ is the following 2X1 vector

$$\underset{\sim}{c}[\underset{\sim}{\sigma}_{(k)}, \underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(k)})] = \begin{bmatrix} \sum_{j=1}^{I}\left\{\frac{J_j[y_{j.}-\alpha(\sigma_{(k)})]^2}{[\sigma_0^{(k)}+J_j\sigma_1^{(k)}]^2}+\frac{\sum_{k=1}^{J_j}(y_{jk}-y_{j.})^2}{[\sigma_0^{(k)}]^2}\right\} \\ \sum_{j=1}^{I}\left\{\frac{J_j[y_{j.}-\alpha(\sigma_{(k)})]^2}{[\sigma_0^{(k)}+J_j\sigma_1^{(k)}]^2}\right\} \end{bmatrix},$$

and where

$$\alpha(\sigma_{(k)}) = \frac{\sum_{j=1}^{I}\left\{\frac{J_jy_{j.}}{[\sigma_0^{(k)}+J_j\sigma_1^{(k)}]}\right\}}{\sum_{j=1}^{I}\left\{\frac{J_j}{[\sigma_0^{(k)}+J_j\sigma_1^{(k)}]}\right\}} .$$

These equations are iterated until convergence is obtained. It is easily seen that it would be difficult to perform these iterations by hand but easy to program them on a computer. A computer program which

handles the completely general case is described in Appendix C.

6.5. Example of a Sequence of Designs not Satisfying Assumption 4.2.4

Recall that Assumption 4.2.4 states that for every i and every $j \in S_s$, $j \neq i$ where $i \in S_s$, there exist two nonnegative constants R_1 and R_2 both less than or equal to one such that

$$\sum_{\ell=1}^{m_j} \left(\frac{\underset{\sim}{u}_\ell^{(j)\prime} \underset{\sim}{u}_k^{(i)}}{\underset{\sim}{u}_k^{(i)\prime} \underset{\sim}{u}_k^{(i)}} \right)^2 \leq R_2,$$

for all but $R_1 m_i$ values of k in the set $\{1,2,\ldots,m_i\}$, where $\underset{\sim}{u}_\ell^{(j)}$ is the ℓ^{th} column of $\underset{\sim}{U}_j$ and $\underset{\sim}{u}_k^{(i)}$ is the k^{th} column of $\underset{\sim}{U}_i$. Furthermore, R_1 and R_2 are such that

$$R_1 + (1 - R_1)\, R_2 \leq \frac{1}{N(S_s)+1} ,$$

where $N(S_s)$ is the number of indices in the set S_s.

The following is an example of a sequence of designs ruled out by this assumption. It is given to illustrate that design sequences

eliminated by this assumption are not design sequences that would be interesting in any event. Let $p_1 = 2$ and let $\underset{\sim}{U}_1$, $n \times m_1$, be any legitimate design matrix (i.e. exactly one 1 in each row) with at least two 1's in each column. Construct $\underset{\sim}{U}_2$ $n \times (m_1+1)$ as follows. For each $j=1,2,\ldots,m_1$ the $(j+1)^{th}$ column of $\underset{\sim}{U}_2$ is just like the j^{th} column of $\underset{\sim}{U}_1$ with one exception. The last 1 in the column is made a zero and that 1 is placed in the same row in column 1 of $\underset{\sim}{U}_2$. For instance, a typical $\underset{\sim}{U}_1$ and $\underset{\sim}{U}_2$ might be

$$\underset{\sim}{U}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \qquad \underset{\sim}{U}_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} .$$

Now since $m_2 = m_1+1$, m_1 and m_2 clearly have the same order of magnitude if a sequence of $\underset{\sim}{U}_1$'s and corresponding $\underset{\sim}{U}_2$'s constructed by the above method is considered. Furthermore $\nu_1 = m_1-1$ and $\nu_2 = m_1 = m_2-1$ by the method of construction. That is, $\underset{\sim}{U}_2$ has been constructed so that the columns of $\underset{\sim}{U}_1$ and $\underset{\sim}{U}_2$ have no linear dependencies except the one forced by the constraint that there is one 1 in each row. Thus Assumption 4.2.3 is satisfied. However, Assumption 4.2.4 is not. Let $\underset{\sim}{U}_1'\underset{\sim}{U}_1 = \underset{\sim}{D}_1$. Then

$$\underset{\sim}{u}_k^{(1)\prime}\, \underset{\sim}{u}_k^{(1)} = d_k^{(1)}$$

and

$$\underset{\sim}{u}_1^{(2)\prime}\,\underset{\sim}{u}_k^{(1)} = 1,\ \underset{\sim}{u}_k^{(2)\prime}\,\underset{\sim}{u}_k^{(1)} = d_k^{(1)} - 1,$$

$$\underset{\sim}{u}_\ell^{(2)\prime}\,\underset{\sim}{u}_k^{(1)} = 0,\ \ell=2,3,\ldots,m_2,\ \ell \neq k\ .$$

But then for <u>all</u> $k=1,2,\ldots,m_1$

$$\sum_{\ell=1}^{m_2}\left(\frac{\underset{\sim}{u}_\ell^{(2)\prime}\,\underset{\sim}{u}_k^{(1)}}{\underset{\sim}{u}_k^{(1)\prime}\,\underset{\sim}{u}_k^{(1)}}\right)^2 = \left[\frac{1}{d_k^{(1)}}\right]^2 + \left[\frac{d_k^{(1)}-1}{d_k^{(1)}}\right]^2 ,$$

$$= \frac{1+[d_k^{(1)}]^2 - 2d_k^{(1)}+1}{[d_k^{(1)}]^2} ,$$

$$= 1 - \frac{2}{d_k^{(1)}} + \frac{2}{[d_k^{(1)}]^2} .$$

This is a monotonically increasing function for $d_k^{(1)} \geq 2$, so its minimum occurs for $d_k^{(1)} = 2$ and gives the bound (since $d_k^{(1)} \geq 2$ for all $k=1,2,\ldots,m_1$ by definition of this problem)

$$\sum_{\ell=1}^{m_2}\left(\frac{\underset{\sim}{u}_\ell^{(2)\prime}\,\underset{\sim}{u}_k^{(1)}}{\underset{\sim}{u}_k^{(1)\prime}\,\underset{\sim}{u}_k^{(1)}}\right)^2 \geq 1 - \frac{2}{2} + \frac{2}{2^2} = \frac{1}{2} ,\quad k=1,2,\ldots,m_1 .$$

But Assumption 4.2.4 demands that at least for a proportion of the k, that the quantity in question be less than or equal to R_2. Clearly $R_2 > \frac{1}{2}$. But then it is impossible to choose any nonnegative R_1 less

than 1 such that

$$R_1 + (1-R_1)R_2 \le \frac{1}{N(S_1)+1} = \frac{1}{2+1} = \frac{1}{3} \quad ,$$

because

$$R_1 + (1-R_1)R_2 \ge R_1 + (1-R_1)\tfrac{1}{2}$$

$$= \tfrac{1}{2} R_1 + \tfrac{1}{2}$$

$$> \frac{1}{3} \; .$$

Thus Assumption 4.2.4 cannot be satisfied for such a design sequence. But such a sequence of designs would never be of interest in any case since almost all treatment combinations overlap. In general, cases that get ruled out by this Assumption 4.2.4 are not of practical interest. All reasonable experimental design sequences pass this assumption.

Another interesting fact is that some sort of assumption like Assumption 4.2.4 is necessary for Theorem 4.4.1. This is true because it can be shown that when $\underset{\sim}{U}_1$ above is a standard design matrix for the row effects in a one-way balanced layout where the number of ones in each column goes to infinity, then the matrix $\underset{\sim}{J}$ needed for Theorem 4.4.1 is singular. But, as shown above, Assumption 4.2.3 will still hold for this case. Thus some stronger assumption than Assumption 4.2.3 is needed for Theorem 4.4.1. Assumption 4.2.4 is such an assumption and as seen above does not rule out any design sequences of consequence.

CHAPTER 7

SUMMARY

7.1. Summary

The results of this paper concerning maximum likelihood estimates in the mixed model of the analysis of variance can be divided into two areas--asymptotic theory and computational procedures. It was proved that under quite general assumptions, the maximum likelihood estimates were consistent and asymptotically efficient in the sense of attaining the Cramér-Rao lower bound for the covariance matrix. A computational procedure was proposed and this procedure performed well in Monte Carlo studies.

In considering asymptotic properties for this model it was necessary to depart from the usual method of proof of such properties. The basic outlines of the proofs (Theorems 3.3.1 and 4.4.1) look similar to proofs used in other situations. The critical difference is that in the model considered here, it is necessary that the sequence of estimates estimating each parameter be allowed to have a separate normalizing sequence. This necessity is most easily recalled by noting that even in simple balanced models, the sums of squares have degrees of freedom which may be of different orders of magnitude. Having a different normalizing sequence for each parameter causes many problems if an attempt is made to push through the new formulation into the old proofs. The Wald-Wolfowitz proof of consistency breaks down in several

places. The Cramér proof, which is used here does not go through directly. However, by the artifice described in Section 4.3--building the normalizing sequence into the parameter--it is possible to restate the Taylor Series expansion type proof in a way that makes sense and is provable.

The idea of different normalizing sequences is the major extension of previous work in this paper. It allows previous results concerning the analysis of variance to be extended by allowing assumptions to be modified. (The assumptions in this paper have been shown to be restrictive only in the sense of eliminating sequences of designs that would not have been of interest in any case.) The rest of the work on asymptotic theory essentially consists of proving the conditions of Theorem 4.4.1 by brute force. An important development contained deep in the detail sections is the idea of separating the length of $\underset{\sim}{y}$ into components relating to linear spaces generated by the $\underset{\sim}{U}$ matrices and then "peeling off" only as much $\underset{\sim}{y}$ (and hence probability) as can be taken care of by the normalizing constant that is available. This is done in Appendix A where the partitioning is defined in Section A.2 and its use is remarked on prior to Proposition A.3.5.

The computational procedure called The Iterative Procedure, which was proposed in Section 5.4, can be motivated in three different ways. Anderson (1971b), who originally proposed it, considered it as an analogy to a certain least squares problem. It has been pointed out here that the problem can be posed as one of functional iteration.

J. N. K. Rao (1973) pointed out that the method was in effect the method of scoring. Together these motivations cause one to have a good deal of faith in the procedure, but the final verification of a procedure's standing should be a proof or demonstration that the good properties mentioned in Section 5.4--guaranteed convergence to a solution of the likelihood equations--hold. It has not been proved here that convergence is sure but in Monte Carlo studies The Iterative Procedure converged over 99% of the time and when it converged it always converged to a solution of the likelihood equations. It was also proved that in a case where closed form solutions to the likelihood equations exist, the procedure will converge to them in one iteration from any initial guess. It can be said that even though proofs of desirable properties do not yet exist, there is ample evidence to conclude that The Iterative Procedure is a very good method for the solution of the likelihood equations in this problem.

This paper would not be complete without mentioning something about the rationale for using maximum likelihood in the analysis of variance. Hartley and Rao (1967) advanced five reasons which were paraphrased in Chapter 2. These reasons are validated even more by the work in this paper. The easy solutions by computer referred to by Hartley and Rao have been made even easier by the addition of The Iterative Procedure, the large sample optimality they cited has been expanded to cover many more cases, and the other reasons remain valid. Thus as a unified theory to cover all cases of balanced or unbalanced models, maximum likelihood has much to recommend it. In the case of balanced models,

it has been shown for several in Chapter 6 that the usual analysis of variance estimates are asymptotically equivalent to the maximum likelihood estimates. This seems to be a general rule. Therefore, it seems that in these balanced models the usual estimates can be used since they do have some small sample optimality properties as well as sharing the large sample optimality properties; also the average user is more familiar with these techniques. The use of the balanced case is to show how well maximum likelihood works there (as in the Monte Carlo studies reported in Section 5.8) to give confidence in its use in the unbalanced case where the usual estimates are very difficult to calculate. In summary, maximum likelihood is a good method to use for the mixed model of the analysis of variance in any situation but is especially important for practical use in the unbalanced layouts where other methods are very difficult to apply.

7.2. Subjects for Further Research

There are several aspects of the topics covered in this paper which may prove fruitful areas for further research. One is an attempt to prove the good properties of The Iterative Procedure. This may in

fact be impossible; every attempt thus far has been stymied by the very difficult arrangement of the nonlinear likelihood equations. However, if such proofs are possible, The Iterative Procedure will certainly be the leading candidate for the best algorithm to compute the maximum likelihood estimates. Another area for further research concerns asymptotic theory. Some research on just how large a design should be for the good asymptotic properties to be approximately true would be very helpful. This might be done in a theorem like those used for the ordinary central limit theorem. More probably it will have to be done with Monte Carlo studies.

Another area where further research would be of practical importance is the area of likelihood ratio tests. Anderson (1969) showed that likelihood ratio tests are easy to derive and that the criterion is just the ratio of determinants of $\underset{\sim}{\Sigma}$ matrices computed under different models. Hartley and Rao (1967) also note that likelihood ratio tests can be used especially easily to test the hypothesis that some of the σ_i's are zero, since the reduced model is just another model of the same form. Both Anderson and Hartley and Rao point out that the likelihood ratio test criterion should be asymptotically distributed as χ^2 random variable under the null hypothesis. However, Anderson's results are for the case where the entire experiment is replicated and Hartley and Rao's results depend on their asymptotic theory with its restricted assumptions. Neither set of results can be extended to the general case using the asymptotic results of this paper because the assumption here is that all σ_i are positive; this assumption of positive σ_i is

critical at several steps of the proof. Under the null hypothesis some of the σ_i equal zero so the results of this paper do not apply. (In fact, the asymptotic distributions in certain simple models are definitely not normal.) Further research which will deal with the asymptotic behavior when some of the σ_i are zero will be very useful, both for its own sake and in dealing with likelihood ratio tests.

APPENDIX A

DETAILS FROM CHAPTER 4

A.1. Definition of a Certain Orthogonal Matrix

R^n can be partitioned into orthogonal subspaces as follows. Let $\mathcal{V}_0$ be such that $R^n = \mathcal{V}_0 \oplus \mathcal{L}(\underset{\sim}{U}_1:\ldots:\underset{\sim}{U}_{p_1})$ and $\mathcal{V}_0$ is orthogonal to $\mathcal{L}(\underset{\sim}{U}_1:\ldots:\underset{\sim}{U}_{p_1})$. For s=1,2,...,c-1 let $\mathcal{V}_s$ be such that $\mathcal{L}(\underset{\sim}{U}_{i_s}:\ldots:\underset{\sim}{U}_{p_1}) = \mathcal{V}_s \oplus \mathcal{L}(\underset{\sim}{U}_{i_{s+1}}:\ldots:\underset{\sim}{U}_{p_1})$ and $\mathcal{V}_s$ is orthogonal to $\mathcal{L}(\underset{\sim}{U}_{i_{s+1}}:\ldots:\underset{\sim}{U}_{p_1})$. Let $\underset{\sim}{\mathcal{V}}_c = \mathcal{L}(\underset{\sim}{U}_{i_c}:\ldots:\underset{\sim}{U}_{p_1})$. (The $\underset{\sim}{U}_i$ matrices are as defined in Section 1.3 and the partition of $\{1,2,\ldots,p_1\}$ into sets S_s is as described in Section 4.2.) Then there are c+1 mutually orthogonal vector spaces such that $R_n = \mathcal{V}_0 \oplus \mathcal{V}_1 \oplus \ldots \oplus \mathcal{V}_c$. Let the dimension of $\mathcal{V}_s$ be $\tilde{m}_s$ and let $\underset{\sim}{H}_s$ be an orthonormal basis for $\mathcal{V}_s$. Then

$$\underset{\sim}{H}'_{s_1}\underset{\sim}{H}_{s_2} = \begin{cases} \underset{\sim}{0} & s_1 \neq s_2 \\ \underset{\sim}{I}_{\tilde{m}_{s_1}} & s_1 = s_2 \end{cases}$$

Thus $\underset{\sim}{P} \equiv [\underset{\sim}{H}_0:\underset{\sim}{H}_1:\ldots:\underset{\sim}{H}_c]$ is an nxn orthogonal matrix; that is, $\underset{\sim}{P}'\underset{\sim}{P} = \underset{\sim}{I} = \underset{\sim}{P}\underset{\sim}{P}'$. Furthermore, for any s=0,1,...,c and any $i \in S^*_{s+1}$, $\underset{\sim}{U}'_i\underset{\sim}{H}_s = \underset{\sim}{0}$ because the columns of $\underset{\sim}{H}_s$ are orthogonal to all vectors in $\mathcal{L}(\underset{\sim}{U}_{i_{s+1}}:\underset{\sim}{U}_{i_{s+1}+1}:\ldots:\underset{\sim}{U}_{p_1})$. This yields

$$\left(\sum_{i=0}^{p_1} b_i \underset{\sim}{G}_i\right)\underset{\sim}{H}_s = \sum_{i=0}^{p_1} b_i \underset{\sim}{U}_i \underset{\sim}{U}_i' \underset{\sim}{H}_s$$

(Let $\underset{\sim}{U}_0 \equiv \underset{\sim}{I}_n$.)

$$= \sum_{i=0}^{i_{s+1}-1} b_i \underset{\sim}{U}_i \underset{\sim}{U}_i' \underset{\sim}{H}_s$$

$$= \sum_{i \notin S^*_{s+1}} b_i \underset{\sim}{G}_i \underset{\sim}{H}_s \;.$$

Now since $\underset{\sim}{T}_0 \underset{\sim}{\Sigma}_0 = \Sigma_0 = \sum_{i=0}^{p_1} \sigma_{0i} \underset{\sim}{G}_i$ is positive definite, there exists a lower triangular matrix $\underset{\sim}{A}$ such that $\underset{\sim}{\Sigma}_0 = \underset{\sim}{A}\underset{\sim}{A}'$. But $\underset{\sim}{P}'\underset{\sim}{\Sigma}_0\underset{\sim}{P}$ is also positive definite and hence there exists $\underset{\sim}{\tilde{T}}$ upper triangular such that $\underset{\sim}{P}'\underset{\sim}{\Sigma}_0\underset{\sim}{P}=\underset{\sim}{\tilde{T}}'\underset{\sim}{\tilde{T}}$. Then if $\underset{\sim}{T} = \underset{\sim}{\tilde{T}}^{-1}$, $\underset{\sim}{T}$ is also upper triangular and $\underset{\sim}{Q} \equiv \underset{\sim}{A}'\underset{\sim}{P}\underset{\sim}{T}$ is an nxn matrix with the following properties.

1) $\underset{\sim}{Q}'\underset{\sim}{Q} = \underset{\sim}{T}'\underset{\sim}{P}'\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{P}\underset{\sim}{T}$,

$$= \underset{\sim}{T}'\underset{\sim}{P}'\underset{\sim}{\Sigma}_0\underset{\sim}{P}\underset{\sim}{T}$$

$$= (\underset{\sim}{\tilde{T}}')^{-1}\underset{\sim}{\tilde{T}}'\,\underset{\sim}{\tilde{T}}\underset{\sim}{\tilde{T}}^{-1}$$

$$= \underset{\sim}{I};$$

that is, $\underset{\sim}{Q}$ is orthogonal.

$$2)\quad \underset{\sim}{Q} = \underset{\sim}{A}'[\underset{\sim}{H}_0:\underset{\sim}{H}_1:\ldots:\underset{\sim}{H}_c]\begin{bmatrix} \underset{\sim}{T}_{00}:\underset{\sim}{T}_{01}:\ldots:\underset{\sim}{T}_{0c} \\ \underset{\sim}{0} \;:\underset{\sim}{T}_{11}:\ldots:\underset{\sim}{T}_{1c} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ \underset{\sim}{0} \;:\; \underset{\sim}{0} \qquad \underset{\sim}{T}_{cc} \end{bmatrix}$$

$$\equiv \underset{\sim}{A}'[\underset{\sim}{H}^*_0:\underset{\sim}{H}^*_1:\ldots:\underset{\sim}{H}^*_c]$$

$$\equiv [\underset{\sim}{Q}_0:\underset{\sim}{Q}_1:\ldots:\underset{\sim}{Q}_c]$$

where $\underset{\sim}{Q}_s = \underset{\sim}{A}'\underset{\sim}{H}^*_s = \underset{\sim}{A}' \sum_{t=0}^{s} \underset{\sim}{H}_t\underset{\sim}{T}_{ts}$ and everything is partitioned so that all multiplications can be properly carried out.

Then it is true that

$$\left(\sum_{i=0}^{p_1} b_i\underset{\sim}{G}_i\right)\underset{\sim}{H}^*_s = \left(\sum_{i=0}^{p_1} b_i\underset{\sim}{U}_i\underset{\sim}{U}'_i\right)\left(\sum_{t=0}^{s} \underset{\sim}{H}_t\underset{\sim}{T}_{ts}\right)$$

$$= \left(\sum_{i=0}^{i_{s+1}-1} b_i\underset{\sim}{U}_i\underset{\sim}{U}'_i\right)\left(\sum_{t=0}^{s} \underset{\sim}{H}_t\underset{\sim}{T}_{ts}\right)$$

$$= \left(\sum_{i \notin S^*_{s+1}} b_i\underset{\sim}{G}_i\right)\underset{\sim}{H}^*_s$$

because H^*_s only involves H_t for $t \le s$. (For $t \le s$, $i \in S^*_{s+1}$ implies $i \in S^*_{t+1}$; thus $\underset{\sim}{G}_i\underset{\sim}{H}_t = \underset{\sim}{0}$. Therefore $\underset{\sim}{G}_i\underset{\sim}{H}^*_s = \underset{\sim}{0}$ for $i \in S^*_{s+1}$.)

3) $\underset{\sim}{Q}$ is orthogonal so $\underset{\sim}{Q}'\underset{\sim}{Q} = \underset{\sim}{I}$. This implies

$$\underset{\sim}{I} = \begin{bmatrix} \underset{\sim}{Q}_0' \\ \cdots \\ \underset{\sim}{Q}_1' \\ \cdots \\ \vdots \\ \cdots \\ \underset{\sim}{Q}_c' \end{bmatrix} [\underset{\sim}{Q}_0 : \underset{\sim}{Q}_1 : \cdots : \underset{\sim}{Q}_c]$$

Therefore

$$\underset{\sim}{Q}_s'\underset{\sim}{Q}_t = \begin{cases} \underset{\sim}{0} & s \neq t \\ \underset{\sim}{I}_{\tilde{m}_s} & s = t \end{cases} .$$

A.2. Definition of Condition A.2.1.

Let $\underset{\sim}{Q} = [\underset{\sim}{Q}_0 : \underset{\sim}{Q}_1 : \cdots : \underset{\sim}{Q}_c]$ be defined as in Section A.1, where each $\underset{\sim}{Q}_s$ is $n \times \tilde{m}_s$. Then if $\underset{\sim}{\Sigma}_0 = \underset{\sim}{A}\underset{\sim}{A}'$ as usual, $\underset{\sim}{z} = \underset{\sim}{A}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha}_0) \sim \eta_n(\underset{\sim}{0}, \underset{\sim}{I}_n)$ whenever $\underset{\sim}{y} \sim \eta_n(\underset{\sim}{X}\underset{\sim}{\alpha}_0, \underset{\sim}{\Sigma}_0)$. If $\underset{\sim}{w}$ is defined by $\underset{\sim}{w} = \underset{\sim}{Q}'\underset{\sim}{z}$, then $\underset{\sim}{w} \sim \eta_n(\underset{\sim}{0}, \underset{\sim}{I}_n)$ also. But $\underset{\sim}{w}$ can be written $\underset{\sim}{w} = [\underset{\sim}{w}_0', \underset{\sim}{w}_1', \ldots, \underset{\sim}{w}_c']'$, where

$\underset{\sim}{w}_s = \underset{\sim}{Q}_s' \underset{\sim}{z} = \underset{\sim}{Q}_s' \underset{\sim}{A}^{-1}(\underset{\sim}{y} - X\underset{\sim}{\alpha}_0)$. Each $\underset{\sim}{w}_s$ is $\tilde{m}_s \times 1$ and $\underset{\sim}{w}_s \sim \underset{\sim}{\eta}_{\tilde{m}_s}(\underset{\sim}{0}, \underset{\sim}{I}_{\tilde{m}_s})$ for $s=0,1,\ldots,c$. Condition A.2.1 is defined as follows:

CONDITION A.2.1. <u>For $\underset{\sim}{w}_s$ as defined above,</u> $\dfrac{\underset{\sim}{w}_s'\underset{\sim}{w}_s}{\tilde{m}_s} \leq \dfrac{11}{10}$, $s=0,1,\ldots,c$.

The following proposition concerning Condition A.2.1 is true.

PROPOSITION A.2.1. Under Assumptions 1.3.1-1.3.6 and 4.2.1-4.2.5, P{<u>Condition A.2.1 is true</u>} $\to 1$ as $n \to \infty$.

PROOF.

P{Condition A.2.1 is not true}

$$= P\left\{ \frac{\underset{\sim}{w}_s'\underset{\sim}{w}_s}{\tilde{m}_s} > \frac{11}{10} \text{ for some } s=0,1,\ldots,c \right\}$$

$$\leq \sum_{s=0}^{c} P\left\{ \frac{\underset{\sim}{w}_s'\underset{\sim}{w}_s}{\tilde{m}_s} > \frac{11}{10} \right\}$$

$$= \sum_{s=0}^{c} P\left\{ \frac{\underset{\sim}{w}_s'\underset{\sim}{w}_s - \tilde{m}_s}{\tilde{m}_s} > \frac{1}{10} \right\}$$

$$\leq \sum_{s=0}^{c} P\left\{ \left| \frac{\underset{\sim}{w}_s'\underset{\sim}{w}_s - \tilde{m}_s}{\tilde{m}_s} \right| > \frac{1}{10} \right\} .$$

But $\underset{\sim}{w}_s'\underset{\sim}{w}_s \sim \chi^2_{\tilde{m}_s}$ and hence $\mathcal{E}(\underset{\sim}{w}_s'\underset{\sim}{w}_s) = \tilde{m}_s$ and $\mathrm{Var}(\underset{\sim}{w}_s'\underset{\sim}{w}_s) = 2\tilde{m}_s$ for $s=0,1,\ldots,c$,

so by a simple application of Chebychev's Inequality

$$P\{\text{Condition A.2.1 is not true}\} \le \sum_{s=0}^{c} \frac{200}{\tilde{m}_s}.$$ But each $\tilde{m}_s \ge \nu_i$ for some $i \in S_s$ where ν_i is as defined in Section 4.2. Thus each $\tilde{m}_s \to \infty$ as $n \to \infty$ by Assumptions 4.2.1 and 4.2.4. This implies that $P\{\text{Condition A.2.1 is not true}\} \to 0$ as $n \to \infty$, which proves the proposition.|||

A.3. Bounds for Various Inner Products and the Characteristic Roots of Various Matrices

In this section bounds for all the terms which must be dealt with in Section A.4 are found. These terms are either traces or inner products. The traces are of the form tr $\underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j$ and tr $\underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j\underset{\sim}{E}_3\underset{\sim}{G}_k$ for certain choices of $\underset{\sim}{E}_1, \underset{\sim}{E}_2$ and $\underset{\sim}{E}_3$; the inner products are of the form $\underset{\sim}{\xi}'\underset{\sim}{X}'\underset{\sim}{E}\underset{\sim}{X}\underset{\sim}{\xi}_2$, $\underset{\sim}{\xi}'\underset{\sim}{X}'\underset{\sim}{E}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$, and $(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'\,\underset{\sim}{E}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$, where $\underset{\sim}{E}$ equals $\underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2$, $\underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j\underset{\sim}{E}_3$ or $\underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j\underset{\sim}{E}_3\underset{\sim}{G}_k\underset{\sim}{E}_4$ for various choices of $\underset{\sim}{E}_1$, $\underset{\sim}{E}_2$, $\underset{\sim}{E}_3$, $\underset{\sim}{E}_4$, $\underset{\sim}{\xi}$, $\underset{\sim}{\xi}_1$ and $\underset{\sim}{\xi}_2$. Bounds will be established for many characterstic roots leading up to bounding the desired inner products and traces.

Let $b > 0$ be given and let $0 < \delta < \frac{b}{2}$. Let $\underset{\sim}{\psi}_{1n} \in S_b(\underset{\sim}{\psi}_{0n})$ and let $\underset{\sim}{\psi}_{2n} \in S_\delta(\underset{\sim}{\psi}_{1n})$. Let $\underset{\sim}{\psi}_{an} = (\beta_a', \tau_{a0}, \tau_{a1}, \ldots, \tau_{ap_1})'$, a=0,1,2 and recall that $\underset{\sim}{\psi}_{1n} \in S_b(\underset{\sim}{\psi}_{0n})$ implies $\|\underset{\sim}{\psi}_{1n} - \underset{\sim}{\psi}_{0n}\| < b$ which in turn implies that

$|\tau_{1i}-\tau_{0i}| < b$ for all $i=0,1,\ldots,p_1$ and $|\beta_{1j}-\beta_{0j}| < b$ for all $j=1,2,\ldots,p_0$. Similarly $|\tau_{1i}-\tau_{2i}| < \delta$ and $|\beta_{1j}-\beta_{2j}| < \delta$. Recall that $\sigma_{0i} = \dfrac{\tau_{0i}}{n_i}$. Let the following condition hold.

CONDITION A.3.1. <u>For a fixed</u> $b > 0$ <u>and with</u> σ_{0i} <u>and</u> n_i <u>defined as in Chapters 1 and 4</u>, $\dfrac{\sigma_{0i}}{2} > \dfrac{b}{n_i}$, $i=0,1,\ldots,p_1$.

It is always possible that Condition A.3.1 holds because the n_i are sequences increasing to infinity and all σ_{0i} are positive. Several consequences follow immediately from the above definitions.

PROPOSITION A.3.1. <u>For</u> $\underset{\sim}{\psi}_{0n}$, $\underset{\sim}{\psi}_{1n}$, $\underset{\sim}{\psi}_{2n}$ <u>as defined above, if Condition A.3.1 is true, then the following statements are true.</u>

$$(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)'(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1) \le p_0 b^2 ,$$

$$(\underset{\sim}{\beta}_1-\underset{\sim}{\beta}_2)'(\underset{\sim}{\beta}_1-\underset{\sim}{\beta}_2) \le p_0 \delta^2 ,$$

$$0 < \frac{\sigma_{0i}}{2} < \frac{\tau_{1i}}{n_i} < \frac{3\sigma_{0i}}{2} ,$$

<u>and</u>

$$0 < \frac{\sigma_{0i}}{4} < \frac{\tau_{2i}}{n_i} < \frac{7\sigma_{0i}}{4} , \quad i=0,1,\ldots,p_1 .$$

PROOF.

$$(\underset{\sim}{\beta}_0 - \underset{\sim}{\beta}_1)'(\underset{\sim}{\beta}_0 - \underset{\sim}{\beta}_1) = \sum_{j=1}^{p_0} (\beta_{0j} - \beta_{1j})^2 \le p_0 b^2 \ .$$

$$(\underset{\sim}{\beta}_1 - \underset{\sim}{\beta}_2)'(\underset{\sim}{\beta}_1 - \underset{\sim}{\beta}_2) = \sum_{j=1}^{p_0} (\beta_{1j} - \beta_{2j})^2 \le p_0 \delta^2 \ .$$

$$\frac{\tau_{1i}}{n_i} = \frac{\tau_{0i}}{n_i} + \frac{\tau_{1i} - \tau_{0i}}{n_i} \ ,$$

$$= \sigma_{0i} + \frac{\tau_{1i} - \tau_{0i}}{n_i} \ ,$$

$$\le \sigma_{0i} + \frac{b}{n_i} < \frac{3\sigma_{0i}}{2} \ ,$$

and

$$\frac{\tau_{1i}}{n_i} \ge \sigma_{0i} - \frac{b}{n_i} > \frac{\sigma_{0i}}{2} > 0 \ .$$

Similarly

$$\frac{\tau_{2i}}{n_i} = \frac{\tau_{0i}}{n_i} + \frac{\tau_{1i} - \tau_{0i}}{n_i} + \frac{\tau_{2i} - \tau_{1i}}{n_i}$$

$$\le \sigma_{0i} + \frac{b}{n_i} + \frac{\delta}{n_i} < \sigma_{0i} + \frac{\sigma_{0i}}{2} + \frac{\sigma_{0i}}{4} = \frac{7\sigma_{0i}}{4} \ ,$$

and

$$\frac{\tau_{2i}}{n_i} \ge \sigma_{0i} - \frac{b}{n_i} - \frac{\delta}{n_i} > \frac{\sigma_{0i}}{4} > 0 \ .$$

|||

Now define $\boldsymbol{T}_a = \sum_{i=0}^{p_1} \frac{\tau_{ai}}{n_1} \boldsymbol{G}_i$, $a=0,1,2$ and recall $\boldsymbol{T}_0 = \boldsymbol{\Sigma}_0$.

Proposition A.3.1 is used to prove the following proposition.

PROPOSITION A.3.2. <u>If Condition A.3.1 is true, the following statements are true.</u>

$$\lambda_{max}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{G}_i) \le \frac{1}{\sigma_{0i}}\,, \quad \lambda_{max}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{T}_1) \le \frac{3}{2}\,, \quad \lambda_{max}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{T}_2) \le \frac{7}{4}\,, \quad \lambda_{max}(\boldsymbol{T}_1^{-1}\boldsymbol{\Sigma}_0) \le 2,$$

$$\lambda_{max}(\boldsymbol{T}_2^{-1}\boldsymbol{\Sigma}_0) \le 4, \quad \lambda_{max}(\boldsymbol{T}_1^{-1}\boldsymbol{T}_2) \le \frac{7}{2}\,, \quad \lambda_{max}(\boldsymbol{T}_2^{-1}\boldsymbol{T}_1) \le 6\,,$$

$$\max_{k=1,2,\ldots,n} |\lambda_k[\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{T}_1-\boldsymbol{T}_2)]| = \max_{k=1,2,\ldots,n} |\lambda_k[\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{T}_2-\boldsymbol{T}_1)]| \le \frac{\delta}{\min\limits_{i=0,1,\ldots,p_1} (n_i\sigma_{0i})}\,,$$

$$\max_{k=1,2,\ldots,n} |\lambda_k[\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{T}_1-\boldsymbol{\Sigma}_0)]| = \max_{k=1,2,\ldots,n} |\lambda_k[\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\Sigma}_0-\boldsymbol{T}_1)]| \le \frac{b}{\min\limits_{i=0,1,\ldots,p_1} (n_i\sigma_{0i})}\,,$$

$$\max_{k=1,2,\ldots,n} |\lambda_1[\boldsymbol{T}_1^{-1}(\boldsymbol{T}_1-\boldsymbol{\Sigma}_0)]| = \max_{k=1,2,\ldots,n} |\lambda_k[\boldsymbol{T}_1^{-1}(\boldsymbol{\Sigma}_0-\boldsymbol{T}_1)]| \le \frac{2b}{\min\limits_{i=0,1,\ldots,p_1} (n_i\sigma_{0i})},$$

$$\max_{k=1,2,\ldots,n} |\lambda_k[\boldsymbol{T}_1^{-1}(\boldsymbol{T}_1-\boldsymbol{T}_2)]| = \max_{k=1,2,\ldots,n} |\lambda_k[\boldsymbol{T}_1^{-1}(\boldsymbol{T}_2-\boldsymbol{T}_1)]| \le \frac{2\delta}{\min\limits_{i=0,1,\ldots,p_1} (n_i\sigma_{0i})}\,,$$

$$\max_{k=1,2,\ldots,n} |\lambda_k[\boldsymbol{T}_2^{-1}(\boldsymbol{T}_1-\boldsymbol{T}_2)]| = \max_{k=1,2,\ldots,n} |\lambda_k[\boldsymbol{T}_2^{-1}(\boldsymbol{T}_2-\boldsymbol{T}_1)]| \le \frac{4\delta}{\min\limits_{i=0,1,\ldots,p_1} (n_i\sigma_{0i})}$$

PROOF.

Continually apply Lemma B.7 to obtain

$$\lambda_{max}(\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i) \le \max(0, \frac{1}{\sigma_{0i}}) = \frac{1}{\sigma_{0i}}$$

since $G_i = \sum_{\substack{j=0 \\ j\neq i}}^{p_1} 0\cdot\underset{\sim}{G}_j + 1\cdot\underset{\sim}{G}_i$.

$$\lambda_{max}(\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{T}_1) \le \max_{i=0,1,\ldots,p_1} \frac{\tau_{1i}/n_i}{\sigma_{0i}} \le \frac{3}{2} .$$

$$\lambda_{max}(\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{T}_2) \le \max_{i=0,1,\ldots,p_1} \frac{\tau_{2i}/n_i}{\sigma_{0i}} \le \frac{7}{4} .$$

$$\lambda_{max}(\underset{\sim}{T}_1^{-1}\underset{\sim}{\Sigma}_0) \le \max_{i=0,1,\ldots,p_1} \frac{\sigma_{0i}}{\tau_{1i}/n_i} \le 2 .$$

$$\lambda_{max}(\underset{\sim}{T}_2^{-1}\underset{\sim}{\Sigma}_0) \le \max_{i=0,1,\ldots,p_1} \frac{\sigma_{0i}}{\tau_{2i}/n_i} \le 4 .$$

$$\lambda_{max}(\underset{\sim}{T}_1^{-1}\underset{\sim}{T}_2) \le \max_{i=0,1,\ldots,p_1} \frac{\tau_{2i}/n_i}{\tau_{1i}/n_i} \le \frac{7}{2} .$$

$$\lambda_{max}(\underset{\sim}{T}_2^{-1}\underset{\sim}{T}_1) \le \max_{i=0,1,\ldots,p_1} \frac{\tau_{1i}/n_i}{\tau_{2i}/n_i} \le 6 .$$

For the remaining inequalities

$$\max_{k=1,2,\ldots,n} \left|\lambda_k[\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)]\right| = \max_{k=1,2,\ldots,n} \left| \frac{\underset{\sim}{x}_k'(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{x}_k}{\underset{\sim}{x}_k'\underset{\sim}{\Sigma}_0\underset{\sim}{x}_k} \right|$$

by Lemma B.6,

$$= \max_{k=1,2,\ldots,n} \left| \frac{\sum_{i=0}^{p_1} \frac{\tau_{1i}-\tau_{2i}}{n_i} \underset{\sim}{x}_k'\underset{\sim}{G}_i\underset{\sim}{x}_k}{\sum_{i=0}^{p_1} \sigma_{0i}\underset{\sim}{x}_k'\underset{\sim}{G}_i\underset{\sim}{x}_k} \right|$$

$$= \max_{k=1,2,\ldots,n} \left| \frac{\sum_{i=0}^{p_1} \frac{\tau_{2i}-\tau_{1i}}{n_i} \underset{\sim}{x}_k'\underset{\sim}{G}_i\underset{\sim}{x}_k}{\sum_{i=0}^{p_1} \sigma_{0i}\underset{\sim}{x}_k'\underset{\sim}{G}_i\underset{\sim}{x}_k} \right|$$

because $\sigma_{0i} > 0$ and $\underset{\sim}{x}_k'\underset{\sim}{G}_i\underset{\sim}{x}_k \geq 0$,

$$= \max_{k=1,2,\ldots,n} \left|\lambda_k[\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{T}_2-\underset{\sim}{T}_1)]\right|$$

$$\leq \max_{k=1,2,\ldots,n} \frac{\sum_{i=0}^{p_1} \frac{|\tau_{1i}-\tau_{2i}|}{n_i} \underset{\sim}{x}_k'\underset{\sim}{G}_i\underset{\sim}{x}_k}{\sum_{i=0}^{p_1} \sigma_{0i}\underset{\sim}{x}_k'\underset{\sim}{G}_i\underset{\sim}{x}_k},$$

$$\leq \max_{i=0,1,\ldots,p_1} \frac{|\tau_{1i}-\tau_{2i}|}{n_i\sigma_{0i}}$$

$$\leq \frac{\delta}{\min\limits_{i=0,1,\ldots,p_1} (n_i \sigma_{0i})}$$

by Lemma B.3. Similarly,

$$\max_{k=1,2,\ldots,n} \left|\lambda_k[\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{T}_1-\underset{\sim}{\Sigma}_0)]\right| \leq \max_{i=0,1,\ldots,p_1} \frac{|\tau_{1i}/n_i-\sigma_{0i}|}{\sigma_{0i}}$$

$$= \max_{i=0,1,\ldots,p_1} \left|\frac{\tau_{1i}-\tau_{0i}}{n_i\sigma_{0i}}\right|$$

$$\leq \frac{b}{\min\limits_{i=0,1,\ldots,p_1} (n_i \sigma_{0i})} .$$

$$\max_{k=1,2,\ldots,n} \left|\lambda_k[\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{\Sigma}_0)]\right| \leq \max_{i=0,1,\ldots,p_1} \frac{|\tau_{1i}/n_i\ -\sigma_{0i}|}{\tau_{1i}/n_i}$$

$$\leq \frac{2b}{\min\limits_{i=0,1,\ldots,p_1} (n_i \sigma_{0i})} .$$

$$\max_{k=1,2,\ldots,n} \left|\lambda_k[\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)]\right| \leq \max_{i=0,1,\ldots,p_1} \frac{|(\tau_{1i}-\tau_{2i})/n_i|}{\tau_{1i}/n_i}$$

$$\leq \max_{i=0,1,\ldots,p_1} \frac{\delta/n_i}{\sigma_{0i}/2}$$

$$\leq \frac{2\delta}{\min_{i=0,1,\ldots,p_1} (n_i\sigma_{0i})} .$$

$$\max_{k=1,2,\ldots,n} |\lambda_k[\underset{\sim}{T}_2^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)]| \leq \max_{i=0,1,\ldots,p_1} \frac{|(\tau_{1i}-\tau_{2i})/n_i|}{\tau_{2i}/n_i}$$

$$\leq \frac{4\delta}{\min_{i=0,1,\ldots,p_1} (n_i\sigma_{0i})} . \quad |||$$

PROPOSITION A.3.3. Let $\underset{\sim}{E}_1, \underset{\sim}{E}_2, \underset{\sim}{E}_3$ and $\underset{\sim}{E}_4$ be any nxn symmetric matrices and $\underset{\sim}{\Sigma}_0 = \underset{\sim}{A}\underset{\sim}{A}'$; then if Condition A.3.1 is true the following statements are true.

$$\lambda_{max}(\underset{\sim}{A}'\underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{E}_2\underset{\sim}{G}_i\underset{\sim}{E}_1\underset{\sim}{A})$$

$$\leq \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_1)|^2 \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_2)|^2 \frac{1}{\sigma_{0i}^2} ,$$

$$\lambda_{max}(\underset{\sim}{A}'\underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j\underset{\sim}{E}_3\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{E}_3\underset{\sim}{G}_j\underset{\sim}{E}_2\underset{\sim}{G}_i\underset{\sim}{E}_1\underset{\sim}{A})$$

$$\leq \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_1)|^2 \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_2)|^2$$

$$\cdot \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_3)|^2 \frac{1}{\sigma_{0i}^2\sigma_{0j}^2}$$

$$\lambda_{\max}(\underset{\sim}{A}'\underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j\underset{\sim}{E}_3\underset{\sim}{G}_k\underset{\sim}{E}_4\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{E}_4\underset{\sim}{G}_k\underset{\sim}{E}_3\underset{\sim}{G}_j\underset{\sim}{E}_2\underset{\sim}{G}_i\underset{\sim}{E}_1\underset{\sim}{A})$$

$$\leq \max_{\ell=1,2,\ldots,n} |\lambda_\ell(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_1)|^2 \max_{\ell=1,2,\ldots,n} |\lambda_\ell(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_2)|^2 \max_{\ell=1,2,\ldots,n} |\lambda_\ell(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_3)|^2$$

$$\cdot \max_{\ell=1,2,\ldots,n} |\lambda_\ell(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_4)|^2 \cdot \frac{1}{\sigma_{0i}^2\sigma_{0j}^2\sigma_{0k}^2} .$$

PROOF.

The proof is given for the first case only; the other cases are proved analogously.

$$\lambda_{\max}(\underset{\sim}{A}'\underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{E}_2\underset{\sim}{G}_i\underset{\sim}{E}_1\underset{\sim}{A})$$

$$= \lambda_{\max}(\underset{\sim}{A}'\underset{\sim}{E}_1\underset{\sim}{A}\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{A}'\underset{\sim}{E}_2\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{E}_2\underset{\sim}{A}\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{A}'\underset{\sim}{E}_1\underset{\sim}{A})$$

$$\leq \lambda_{\max}(\underset{\sim}{A}'\underset{\sim}{E}_1\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{E}_1\underset{\sim}{A})\ \lambda_{\max}(\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t})\ \lambda_{\max}(\underset{\sim}{A}'\underset{\sim}{E}_2\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{E}_2\underset{\sim}{A})$$

by two applications of Lemma B.8 ,

$$= \lambda_{max}(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_1)^2\ \lambda_{max}(\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i)^2\ \lambda_{max}(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_2)^2$$

by Lemma B.9,

$$\leq \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_1)|^2 \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{\Sigma}_0\underset{\sim}{E}_2)|^2 \cdot \frac{1}{\sigma_{0i}^2}$$

by Lemmae B.14 and B.7. |||

Proposition A.3.4 deals with inner products.

PROPOSITION A.3.4. <u>Let $\underset{\sim}{\xi}_1$ and $\underset{\sim}{\xi}_2$ be any p_0x1 vectors and $\underset{\sim}{F}_0$, $\underset{\sim}{F}_1$ and $\underset{\sim}{F}_2$ any nxn matrices; then the following statements are true.</u>

$$(\underset{\sim}{\xi}_1'\underset{\sim}{X}'\underset{\sim}{F}_1\underset{\sim}{X}\underset{\sim}{\xi}_2)^2 \leq (\underset{\sim}{\xi}_1'\underset{\sim}{\xi}_1)(\underset{\sim}{\xi}_2'\underset{\sim}{\xi}_2)\ \lambda_{max}^2(\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X})\ \lambda_{max}(\underset{\sim}{A}'\underset{\sim}{F}_1'\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A}),$$

$$[\underset{\sim}{\xi}_1'\underset{\sim}{X}'\underset{\sim}{F}_1\underset{\sim}{F}_2(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)]^2 \leq (\underset{\sim}{\xi}_1'\underset{\sim}{\xi}_1)\ \lambda_{max}(\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X})\ \lambda_{max}(\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A})$$

$$\cdot\ (\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0),$$

$$[(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'\underset{\sim}{F}_0\underset{\sim}{F}_1\underset{\sim}{F}_2(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)]^2 \leq \lambda_{max}(\underset{\sim}{A}'\underset{\sim}{F}_1'\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A})(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'\underset{\sim}{F}_0'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_0(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$$

$$\cdot\ (\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'\underset{\sim}{F}_2\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2'(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)\ .$$

PROOF.

$$(\underset{\sim}{\xi}_1'\underset{\sim}{X}'\underset{\sim}{F}_1\underset{\sim}{X}\underset{\sim}{\xi}_2)^2 = (\underset{\sim}{\xi}_1'\underset{\sim}{X}'\underset{\sim}{A}^{-t}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A}\underset{\sim}{A}^{-1}\underset{\sim}{X}\underset{\sim}{\xi}_2)^2$$

$$\leq (\underset{\sim}{\xi}_1'\underset{\sim}{X}'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{X}\underset{\sim}{\xi}_1)(\underset{\sim}{\xi}_2'\underset{\sim}{X}'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{X}\underset{\sim}{\xi}_2)\lambda_{max}(\underset{\sim}{A}'\underset{\sim}{F}_1'\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A})$$

by Lemma B.12,

$$\leq (\underset{\sim}{\xi}_1'\underset{\sim}{\xi}_1)(\underset{\sim}{\xi}_2'\underset{\sim}{\xi}_2)\lambda^2_{max}(\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X})\lambda_{max}(\underset{\sim}{A}'\underset{\sim}{F}_1'\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A})$$

by definition of characteristic root. The other cases are proved analogously. |||

Now $\underset{\sim}{A}'\underset{\sim}{F}_1'\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A}$ of Proposition A.3.4 will be of the correct form to plug into Proposition A.3.3. It remains to bound terms of the form $(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$. This is done using Condition A.2.1. Let $\underset{\sim}{Q}$ be defined as in Section A.1 and let $\underset{\sim}{w}$ be as defined in Section A.2. Then

$$(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0) = \underset{\sim}{A}\underset{\sim}{A}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$$

$$= \underset{\sim}{A}\underset{\sim}{Q}\underset{\sim}{Q}'\underset{\sim}{A}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$$

$$= \underset{\sim}{A}\underset{\sim}{Q}\underset{\sim}{w}$$

$$= \underset{\sim}{A}[\underset{\sim}{Q}_0 : \underset{\sim}{Q}_1 : \ldots : \underset{\sim}{Q}_c] \begin{bmatrix} \underset{\sim}{w}_0 \\ \underset{\sim}{w}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{w}_c \end{bmatrix}$$

$$= \underset{\sim}{A} \sum_{s=0}^{c} \underset{\sim}{Q}_s \underset{\sim}{w}_s .$$

This yields

$$(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha}_0)' \underset{\sim}{F}_2' \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{F}_2 (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha}_0)$$

$$= \sum_{t=0}^{c} \sum_{s=0}^{c} \underset{\sim}{w}_t' \underset{\sim}{Q}_t' \underset{\sim}{A}' \underset{\sim}{F}_2' \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{F}_2 \underset{\sim}{A} \underset{\sim}{Q}_s \underset{\sim}{w}_s .$$

The Cauchy-Schwarz Inequality gives a bound for each term as

$$(\underset{\sim}{w}_t' \underset{\sim}{Q}_t' \underset{\sim}{A}' \underset{\sim}{F}_2' \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{F}_2 \underset{\sim}{A} \underset{\sim}{Q}_s \underset{\sim}{w}_s)^2$$

$$\leq (\underset{\sim}{w}_t' \underset{\sim}{Q}_t' \underset{\sim}{A}' \underset{\sim}{F}_2' \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{F}_2 \underset{\sim}{A} \underset{\sim}{Q}_t \underset{\sim}{w}_t)(\underset{\sim}{w}_s' \underset{\sim}{Q}_s' \underset{\sim}{A}' \underset{\sim}{F}_2' \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{F}_2 \underset{\sim}{A} \underset{\sim}{Q}_s \underset{\sim}{w}_s).$$

But

$$\underset{\sim}{w}_s' \underset{\sim}{Q}_s' \underset{\sim}{A}' \underset{\sim}{F}_2' \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{F}_2 \underset{\sim}{A} \underset{\sim}{Q}_s \underset{\sim}{w}_s \leq \underset{\sim}{w}_s' \underset{\sim}{w}_s \, \lambda_{max}(\underset{\sim}{Q}_s' \underset{\sim}{A}' \underset{\sim}{F}_2' \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{F}_2 \underset{\sim}{A} \underset{\sim}{Q}_s)$$

and $\underset{\sim}{w}_s' \underset{\sim}{w}_s \leq \frac{11}{10} \tilde{m}_s$ by Condition A.2.1. Thus bounds are required on

$\lambda_{max}(\underset{\sim}{Q}_s'\underset{\sim}{A}'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2\underset{\sim}{A}\underset{\sim}{Q}_s)$ for various choices of $\underset{\sim}{F}_2$. These are provided by the next three propositions.

PROPOSITION A.3.5. <u>If</u> $\underset{\sim}{F}_2 = \underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}$ <u>and if Condition A.3.1 is true then the following statements are true.</u>

<u>If</u> $i \notin S^*_{s+1}$ $\lambda_{max}(\underset{\sim}{Q}_s'\underset{\sim}{A}'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2\underset{\sim}{A}\underset{\sim}{Q}_s) \leq \dfrac{1}{\sigma^2_{0i}}$.

<u>If</u> $i \in S^*_{s+1}$ $\lambda_{max}(\underset{\sim}{Q}_s'\underset{\sim}{A}'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2\underset{\sim}{A}\underset{\sim}{Q}_s) = 0$.

PROOF.

Recall $S^*_{s+1} = \{i_{s+1}, i_{s+1}+1, \ldots, p_1\}$. Then

$$\lambda_{max}(\underset{\sim}{Q}_s'\underset{\sim}{A}'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2\underset{\sim}{A}\underset{\sim}{Q}_s) = \lambda_{max}(\underset{\sim}{Q}_s'\underset{\sim}{A}'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{A}\underset{\sim}{Q}_s)$$

$$= \lambda_{max}(\underset{\sim}{Q}_s'\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{Q}_s)$$

$$= \sup_{\underset{\sim}{\gamma} \neq \underset{\sim}{0}} \frac{\underset{\sim}{\gamma}'\underset{\sim}{Q}_s'\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{Q}_s\underset{\sim}{\gamma}}{\underset{\sim}{\gamma}'\underset{\sim}{\gamma}}$$

$$= \sup_{\underset{\sim}{\gamma} \neq \underset{\sim}{0}} \frac{\underset{\sim}{\gamma}'\underset{\sim}{Q}_s'\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{Q}_s\underset{\sim}{\gamma}}{\underset{\sim}{\gamma}'\underset{\sim}{Q}_s'\underset{\sim}{Q}_s\underset{\sim}{\gamma}}$$

because $\underset{\sim}{Q}_s'\underset{\sim}{Q}_s = \underset{\sim}{I}$,

$$\leq \sup_{\underset{\sim}{x}\neq\underset{\sim}{0}} \frac{\underset{\sim}{x}'\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{x}}{\underset{\sim}{x}'\underset{\sim}{x}}$$

$$= \lambda_{max}^2(\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i)$$

$$\leq \frac{1}{\sigma_{0i}^2} .$$

However, if $i \in S_{s+1}^*$ then $\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{Q}_s = \underset{\sim}{G}_i\underset{\sim}{H}_s^* = \underset{\sim}{0}$ as was shown in Section A.1. Then the matrix in question is the zero matrix and hence has characteristic roots all zero. |||

PROPOSITION A.3.6. <u>If $\underset{\sim}{F}_2 = (\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$ where $\underset{\sim}{T}_1$ is as above and if Condition A.3.1 is true then the following statement is true.</u>

$$\lambda_{max}(\underset{\sim}{Q}_s'\underset{\sim}{A}'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2\underset{\sim}{A}\underset{\sim}{Q}_s) \leq \frac{b^2}{\min\limits_{i=0,1,\ldots,i_{s+1}-1} (n_i\sigma_{0i})^2} .$$

PROOF.

$$\lambda_{max}(\underset{\sim}{Q}_s'\underset{\sim}{A}'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2\underset{\sim}{A}\underset{\sim}{Q}_s) = \lambda_{max}(\underset{\sim}{Q}_s'\underset{\sim}{A}^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{A}^{-t}\underset{\sim}{Q}_s).$$

But

$$(\underset{\sim}{T}_1-\underset{\sim}{\Sigma}_0)\underset{\sim}{A}^{-t}\underset{\sim}{Q}_s = \left(\sum_{i=0}^{p_1} \frac{(\tau_{0i}-\tau_{1i})}{n_i}\underset{\sim}{G}_i\right)\underset{\sim}{H}_s^*$$

$$= \left(\sum_{i=0}^{i_{s+1}-1} \frac{(\tau_{0i}-\tau_{1i})}{n_i} \underset{\sim}{G}_i \right) \underset{\sim}{H}_s^*$$

$$= \underset{\sim}{F}^* \underset{\sim}{H}_s^*$$

$$= \underset{\sim}{F}^* \underset{\sim}{A}^{-t} \underset{\sim}{A}' \underset{\sim}{H}_s^*$$

$$= \underset{\sim}{F}^* \underset{\sim}{A}^{-t} \underset{\sim}{Q}_s$$

where $\underset{\sim}{F}^* = \sum_{i=0}^{i_{s+1}-1} \frac{(\tau_{0i}-\tau_{1i})}{n_i} \underset{\sim}{G}_i$, which is symmetric. Therefore

$$\lambda_{max}(\underset{\sim}{Q}_s' \underset{\sim}{A}' \underset{\sim}{F}_2' \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{F}_2 \underset{\sim}{A} \underset{\sim}{Q}_s) = \lambda_{max}(\underset{\sim}{Q}_s' \underset{\sim}{A}^{-1} \underset{\sim}{F}^* \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{F}^* \underset{\sim}{A}^{-t} \underset{\sim}{Q}_s)$$

$$= \sup_{\underset{\sim}{y} \neq \underset{\sim}{0}} \frac{\underset{\sim}{y}' \underset{\sim}{Q}_s' \underset{\sim}{A}^{-1} \underset{\sim}{F}^* \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{F}^* \underset{\sim}{A}^{-t} \underset{\sim}{Q}_s \underset{\sim}{y}}{\underset{\sim}{y}' \underset{\sim}{y}}$$

$$\leq \sup_{\underset{\sim}{x} \neq \underset{\sim}{0}} \frac{\underset{\sim}{x}' \underset{\sim}{A}^{-1} \underset{\sim}{F}^* \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{F}^* \underset{\sim}{A}^{-t} \underset{\sim}{x}}{\underset{\sim}{x}' \underset{\sim}{x}}$$

as in Proposition A.3.5,

$$= \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{F}^*)|^2 .$$

$$\leq \frac{b^2}{\min\limits_{i=0,1,\ldots,i_{s+1}-1} (n_i\sigma_{0i})^2}$$

by Lemma B.7 just as in Proposition A.3.2. |||

PROPOSITION A.3.7. <u>If $\underset{\sim}{F}_2 = (\underset{\sim}{T}_1 - \underset{\sim}{T}_2)\underset{\sim}{\Sigma}_0^{-1}$ where $\underset{\sim}{T}_1$ and $\underset{\sim}{T}_2$ are as above and if Condition A.3.1 is true then the following statement is true.</u>

$$\lambda_{max}(\underset{\sim}{Q}_s'\underset{\sim}{A}'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2\underset{\sim}{A}\underset{\sim}{Q}_s) \leq \frac{\delta^2}{\min\limits_{i=0,1,\ldots,i_{s+1}-1} (n_i\sigma_{0i})^2} .$$

PROOF.

Proceed exactly as in the proof of Proposition A.3.6 but at last step, instead of $|\tau_{0i}-\tau_{1i}| < b$ use $|\tau_{1i}-\tau_{2i}| < \delta$. |||

There are two different types of $\underset{\sim}{F}_2$ which will be encountered. However, the necessary bounds reduce to those of the form above as will be seen. The first different $\underset{\sim}{F}_2$ is $\underset{\sim}{F}_2 = \underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}$. But

$$\underset{\sim}{T}_1^{-1} = \underset{\sim}{\Sigma}_0^{-1} + \underset{\sim}{T}_1^{-1} - \underset{\sim}{\Sigma}_0^{-1}$$

$$= \underset{\sim}{\Sigma}_0^{-1} + \underset{\sim}{\Delta}$$

where $\underset{\sim}{\Delta} = \underset{\sim}{T}_1^{-1} - \underset{\sim}{\Sigma}_0^{-1} = \underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{T}_1^{-1} = \underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$. Therefore $\underset{\sim}{T}_1^{-1}\underset{\sim}{B}\underset{\sim}{T}_1^{-1} = \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{B}\underset{\sim}{\Sigma}_0^{-1} + \underset{\sim}{\Delta}\underset{\sim}{B}\underset{\sim}{\Delta} + \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{B}\underset{\sim}{\Delta} + \underset{\sim}{\Delta}\underset{\sim}{B}\underset{\sim}{\Sigma}_0^{-1}$ for any $\underset{\sim}{B}$. In this case

$B = G_i A^{-t} A^{-1} G_i$ and so

$$\gamma' Q_s A' F_2' A^{-t} A^{-1} F_2 A Q_s \gamma = \gamma' Q_s A' T_1^{-1} G_i A^{-t} A^{-1} G_i T_1^{-1} A Q_s \gamma$$

$$= \gamma' Q_s A \Sigma_0^{-1} G_i A^{-t} A^{-1} G_i \Sigma_0^{-1} A Q_s \gamma$$

$$+ \gamma' Q_s' A' \Sigma_0^{-1} (\Sigma_0 - T_1) T_1^{-1} G_i A^{-t} A^{-1} G_i T_1^{-1} (\Sigma_0 - T_1) \Sigma_0^{-1} A Q_s \gamma$$

$$+ \gamma' Q_s' A' \Sigma_0^{-1} G_i A^{-t} A^{-1} G_i T_1^{-1} (\Sigma_0 - T_1) \Sigma_0^{-1} A Q_s \gamma$$

$$+ \gamma' Q_s' A' \Sigma_0^{-1} (\Sigma_0 - T_1) T_1^{-1} G_i A^{-t} A^{-1} G_i \Sigma_0^{-1} A Q_s \gamma \ .$$

The first term is exactly as in Proposition A.3.5. The second term by the same reasoning as Propositions A.3.6 and A.3.4 is less than

$$\gamma' \gamma \, \lambda^2_{\max}(\Sigma_0^{-1} G_i) \, \lambda^2_{\max}(T_1^{-1} \Sigma_0) \max_{k=1,2,\ldots,n} |\lambda_k(\Sigma_0^{-1} F^*)|^2 \ .$$

The third and fourth terms are equal and by one application of the Cauchy-Schwarz Inequality, their squares are bounded by the product of the first two terms. Thus bounds exist for all terms based on Propositions A.3.5 and A.3.6.

The second different term is $\underset{\sim}{F}_2 = (\underset{\sim}{T}_1 - \underset{\sim}{T}_2)\underset{\sim}{T}_1^{-1}$. This yields a decomposition as above with $\underset{\sim}{B} = (\underset{\sim}{T}_1 - \underset{\sim}{T}_2)\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}(\underset{\sim}{T}_1 - \underset{\sim}{T}_2)$. As above the squares of the third and fourth terms will be bounded by the product of the first two terms. The first term will be $\underset{\sim}{\gamma}'\underset{\sim}{Q}_s'\underset{\sim}{A}'\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{T}_1 - \underset{\sim}{T}_2)\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}(\underset{\sim}{T}_1 - \underset{\sim}{T}_2)\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{A}\underset{\sim}{Q}_s\underset{\sim}{\gamma}$, which is exactly the same term that is bounded in Proposition A.3.7. The second term will be $\underset{\sim}{\gamma}'\underset{\sim}{Q}_s'\underset{\sim}{A}'\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)(\underset{\sim}{T}_1 - \underset{\sim}{T}_2)\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}(\underset{\sim}{T}_1 - \underset{\sim}{T}_2)(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{A}\underset{\sim}{Q}_s\underset{\sim}{\gamma}$; it is easily seen that this term will be bounded by

$$\underset{\sim}{\gamma}'\underset{\sim}{\gamma} \max_{k=1,2,\ldots,n} |\lambda_k[\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{T}_1 - \underset{\sim}{T}_2)]|^2 \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{F}^*)|^2.$$ Thus all terms are bounded by previous propositions.

Bounds have now been found for all necessary inner products; bounds for traces are covered by the next two propositions.

PROPOSITION A.3.8. <u>If $\underset{\sim}{E}_1$, $\underset{\sim}{E}_2$ and $\underset{\sim}{E}_3$ are any nxn matrices, then the following statements are true</u>.

$$|\mathrm{tr}\ \underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j| \le \min(m_i, m_j) \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{A}'\underset{\sim}{E}_1\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{E}_2\underset{\sim}{A})| \cdot \frac{1}{\sigma_{0i}\sigma_{0j}}$$

$$|\mathrm{tr}\ \underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j\underset{\sim}{E}_3\underset{\sim}{G}_k| \le \min(m_i, m_j, m_k) \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{A}'\underset{\sim}{E}_1\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{E}_2\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{E}_3\underset{\sim}{A})| \frac{1}{\sigma_{0i}\sigma_{0j}\sigma_{0k}}.$$

PROOF.

The proof is given for the first case; the second is proved analogously. $\underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j$ has rank at most $\min(m_i, m_j)$ and hence has at

most $\min(m_i, m_j)$ nonzero characteristic roots.

$$|\operatorname{tr} \underset{\sim}{E}_1 \underset{\sim}{G}_i \underset{\sim}{E}_2 \underset{\sim}{G}_j| = |\sum_{k=1}^{n} \lambda_k(\underset{\sim}{E}_1 \underset{\sim}{G}_i \underset{\sim}{E}_2 \underset{\sim}{G}_j)|$$

$$\leq \sum_{k=1}^{n} |\lambda_k(\underset{\sim}{E}_1 \underset{\sim}{G}_i \underset{\sim}{E}_2 \underset{\sim}{G}_j)|$$

$$\leq \min(m_i, m_j) \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{E}_1 \underset{\sim}{G}_i \underset{\sim}{E}_2 \underset{\sim}{G}_j)| \, .$$

But

$$\max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{E}_1 \underset{\sim}{G}_i \underset{\sim}{E}_2 \underset{\sim}{G}_j)| = \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{U}_j' \underset{\sim}{A}^{-t} \underset{\sim}{A}' \underset{\sim}{E}_1 \underset{\sim}{G}_i \underset{\sim}{E}_2 \underset{\sim}{A} \underset{\sim}{A}^{-1} \underset{\sim}{U}_j)|$$

by Lemma B.11 and $\underset{\sim}{G}_j = \underset{\sim}{U}_j \underset{\sim}{U}_j'$,

$$\leq \lambda_{max}(\underset{\sim}{U}_j' \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{U}_j) \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{A}' \underset{\sim}{E}_1 \underset{\sim}{G}_i \underset{\sim}{E}_2 \underset{\sim}{A})|$$

by Lemma B.8,

$$\leq \lambda_{max}(\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j) \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{U}_i' \underset{\sim}{A}^{-t} \underset{\sim}{A}' \underset{\sim}{E}_2 \underset{\sim}{A} \underset{\sim}{A}' \underset{\sim}{E}_1 \underset{\sim}{A} \underset{\sim}{A}^{-1} \underset{\sim}{U}_i)|$$

by Lemma B.11 again,

$$\leq \lambda_{max}(\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j) \lambda_{max}(\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i) \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{A}' \underset{\sim}{E}_2 \underset{\sim}{A} \underset{\sim}{A}' \underset{\sim}{E}_1 \underset{\sim}{A})|$$

by Lemma B.8 again,

$$\leq \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{A}' \underset{\sim}{E}_1 \underset{\sim}{A} \underset{\sim}{A}' \underset{\sim}{E}_2 \underset{\sim}{A})| \frac{1}{\sigma_{0i}\sigma_{0j}}$$

by Proposition A.3.2. |||

The reason that $\lambda_k(\mathbf{A}'\mathbf{E}_1\mathbf{A}\mathbf{A}'\mathbf{E}_2\mathbf{A})$ cannot be simplified further is that in general $\max|\lambda_k(\mathbf{A}\mathbf{B})| \nleq \max|\lambda_k(\mathbf{A})| \max|\lambda_k(\mathbf{B})|$. However, the choices of $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_3$ needed in Chapter 4 are always of a form that allows simplification. In particular, Proposition A.3.9 is true.

PROPOSITION A.3.9. If $\mathbf{T}_1, \mathbf{T}_2, \mathbf{A}$ are as above and if Condition A.3.1 is true then for the following choices of $\mathbf{E}_1$ and $\mathbf{E}_2$, the cited bounds result for $\max\limits_{k=1,2,\ldots,n} |\lambda_k(\mathbf{A}'\mathbf{E}_1\mathbf{A}\mathbf{A}'\mathbf{E}_2\mathbf{A})|$.

$\mathbf{E}_1$	$\mathbf{E}_2$	Bound for $\max|\lambda_k(\mathbf{A}'\mathbf{E}_1\mathbf{A}\mathbf{A}'\mathbf{E}_2\mathbf{A})|$
$\boldsymbol{\Sigma}_0^{-1}$	$\boldsymbol{\Sigma}_0^{-1}$	1
$\boldsymbol{\Sigma}_0^{-1}$	$\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\Sigma}_0-\mathbf{T}_1)\mathbf{T}_1^{-1}$	$\dfrac{2b}{\min\limits_{i=0,1,\ldots,p_1}(n_i\sigma_{0i})}$
$\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\Sigma}_0-\mathbf{T}_1)\mathbf{T}_1^{-1}$	$\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\Sigma}_0-\mathbf{T}_1)T_1^{-1}$	$\dfrac{4b^2}{\min\limits_{i=0,1,\ldots,p_1}(n_i\sigma_{0i})^2}$
$\mathbf{T}_1^{-1}$	$\mathbf{T}_1^{-1}$	4
$\mathbf{T}_1^{-1}$	$\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}$	$\dfrac{16\delta}{\min\limits_{i=0,1,\ldots,p_1}(n_i\sigma_{0i})}$
$\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}$	$\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}$	$\dfrac{64\delta^2}{\min\limits_{i=0,1,\ldots,p_1}(n_i\sigma_{0i})^2}$

<u>The following choices of $\underset{\sim}{E}_1$, $\underset{\sim}{E}_2$ and $\underset{\sim}{E}_3$ yield the cited bounds for</u>

$$\max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{A}'\underset{\sim}{E}_1\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{E}_2\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{E}_3\underset{\sim}{A})| .$$

$\underset{\sim}{E}_1$	$\underset{\sim}{E}_2$	$\underset{\sim}{E}_3$	Bound
$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{T}_1^{-1}$	8
$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\dfrac{32\delta}{\min\limits_{i=0,1,\ldots,p_1}(n_i\sigma_{0i})}$
$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\dfrac{128\delta^2}{\min\limits_{i=0,1,\ldots,p_1}(n_i\sigma_{0i})^2}$
$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\dfrac{512\delta^3}{\min\limits_{i=0,1,\ldots,p_1}(n_i\sigma_{0i})^3}$

PROOF.

The first case is obvious. The second case simplifies to $\max|\lambda_k[\underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)]|$ which is bounded by Proposition A.3.2. In cases three and four Lemma B.14 applies as it does in the sixth case. The fifth case is proved here for illustrative purposes.

$$\max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{A}'\underset{\sim}{T}_1^{-1}\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}\underset{\sim}{A}| = \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{\Sigma}_0\underset{\sim}{T}_1^{-1}\underset{\sim}{\Sigma}_0\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}|$$

by Lemma B.11,

$$\leq \lambda^2_{\max}(\underset{\sim}{\Sigma}_0\underset{\sim}{T}_1^{-1}) \max_{k=1,2,\ldots,n} |\lambda_k[(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}]|$$

by Lemma B.13 and B.14,

$$\leq 4 \cdot \frac{4\delta}{\min(n_i\sigma_{0i})}$$

by Proposition A.3.2. Similar logic applies to the second set of bounds. Both sets of bounds are unchanged upon permutations of $\underset{\sim}{E}_1$, $\underset{\sim}{E}_2$ and $\underset{\sim}{E}_3$. |||

This section concludes with some remarks which firm up the bounds obtained in the rest of the section.

PROPOSITION A.3.10. <u>There exists some constant</u> B <u>such that the following statements are true</u>.

$$\frac{1}{n^2_{p_1+1}} \lambda_{\max}(\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X}) \leq B; \quad \frac{\tilde{m}_s}{\min\limits_{i=0,1,\ldots,i_{s+1}-1} n_i^2} \leq B \text{ and } \frac{\tilde{m}_s}{\min\limits_{i=0,1,\ldots,i_{s+1}-1}(n_i\sigma_{0i})^2} \leq B,$$

$$s=0,1,\ldots,c; \quad \frac{\min(m_i,m_j)}{n_in_j} \leq B \text{ and } \frac{\min(m_i,m_j,m_k)}{n_in_jn_k} \to 0 \quad i,j,k=0,1,\ldots,p_1.$$

PROOF.

Recall $n^2_{p_1+1} = \nu_{p_1+1}$. Assumption 4.2.5 then guarantees the first bound. Since $\tilde{m}_s$ and all the n_i increase to infinity and all the σ_{0i} are finite, the second two terms are either bounded or unbounded together. Now by Assumption 4.2.3, each $\nu_i (=n_i^2)$ is the same order of magnitude as m_i. But $\tilde{m}_s = \dim \mathcal{V}_s$ defined in Section A.1, thus

$\tilde{m}_s = \dim\{\mathcal{L}(\underset{\sim}{U}_{i_s}:\ldots:\underset{\sim}{U}_{p_1})\} - \dim\{\mathcal{L}(\underset{\sim}{U}_{i_{s+1}}:\ldots:\underset{\sim}{U}_{p_1})\}$. Since

$\dim\{\mathcal{L}(\underset{\sim}{U}_{i_s}:\ldots:\underset{\sim}{U}_{p_1})\} \geq m_i$ for any $i \in S_s$ and $\dim\{\mathcal{L}(\underset{\sim}{U}_{i_s}:\ldots:\underset{\sim}{U}_{p_1})\} \leq \sum_{j\in S_s^*} m_j$,

it follows that $m_i - \sum_{j\in S_{s+1}^*} m_j \leq \tilde{m}_s \leq \sum_{j\in S_s^*} m_j$ for any $i \in S_s$. But then

$\frac{\tilde{m}_s}{m_i}$ is bounded by $\sum_{j\in S_s^*} \frac{m_j}{m_i} \to \sum_{j\in S_s} \rho_{ji} < \infty$ and $\frac{m_i - \sum_{j\in S_{s+1}^*} m_j}{m_i} \to 1 - \sum_{j\in S_{s+1}^*} \rho_{ji} = 1$

because $\rho_{ji} = 0$ for $i \in S_s$, $j \in S_{s+1}^*$. Thus each $\tilde{m}_s$ has the same order of magnitude as each m_i for $i \in S_s$. It is thus sufficient to show that $\min_{i=0,1,\ldots,i_{s+1}-1} n_i^2 = n_j$ where $j \in S_s$ at least for n large. However for $t < s$ and $i \in S_s$ and $j \in S_t$, $p_{ji} = \lim_{n\to\infty} \frac{m_j}{m_i} = +\infty$ by definition of the sets S_s. Therefore, the minimum m_i and hence the minimum n_i must eventually equal n_j for some $j \in S_s$. But then the minimum n_i and $\tilde{m}_s$ have the same order of magnitude and the second and third expressions are both bounded.

For the last two statements assume without loss of generality, that $i \geq j \geq k$. Then $i \in S_s$, $j \in S_t$ with $s \geq t$. If s=t, either m_i or m_j could be the minimum but then n_i and n_j both have the same order of magnitude and $n_i n_j$ has the same order of magnitude as

$\min[m_i, m_j] \geq \min[m_i, m_j, m_k]$. Hence the first expression is bounded and since $n_k \to \infty$ the second converges to zero. If $s > t$ then m_i will eventually be the minimum and n_j and n_k are both of greater order of magnitude than n_i. Thus both expressions converge to zero.

Since all expressions are bounded, a common bound for all of them can be chosen. |||

A.4. Lemmae Used to Prove Theorem 4.4.1

This section contains the lemmae required to prove Theorem 4.4.1. Each lemmae is stated and proved in a separate subsection. These lemmae are referred to in the proof of Theorem 4.4.1 given in Section 4.5.

A.4.1. Proof of Conclusion 4.4.1.i--The Positive Definiteness of $\underset{\sim}{J}$

LEMMA A.4.1. The pxp matrix $\underset{\sim}{J}$ defined by

$$[\underset{\sim}{J}]_{ij} = \lim_{n\to\infty} \left[-\mathcal{E}_0 \left(\frac{\partial^2 \lambda(\underset{\sim}{y}, \underset{\sim}{\psi})}{\partial\psi_i \partial\psi_j} \bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} \right) \right], \; i,j=1,2,\ldots,p \text{ is positive}$$

definite.

PROOF.

It was shown in Section 4.3 that $\underset{\sim}{J}=\begin{bmatrix} \underset{\sim}{C}_0 & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{C}_1 \end{bmatrix}$ and that $\underset{\sim}{C}_0$ was positive definite by Assumption 4.2.5. It remains to show that the $(p_1+1)\times(p_1+1)$ matrix $\underset{\sim}{C}_1$ is positive definite, where $[\underset{\sim}{C}_1]_{ij} = \frac{1}{2} \lim_{n\to\infty} \frac{1}{n_i n_j} \operatorname{tr} \underset{\sim}{T}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{T}_0^{-1} \underset{\sim}{G}_j$, $i,j=0,1,\ldots,p_1$.

Let $b_0, b_1, \ldots, b_{p_1}$ be arbitrary constants, not all zero. It is required to show that

$$\sum_{i=0}^{p_1} \sum_{j=0}^{p_1} b_i \, b_j (\underset{\sim}{C}_1)_{ij} > 0.$$

But

$$2 \sum_{i=0}^{p_1} \sum_{j=0}^{p_1} b_i b_j (\underset{\sim}{C}_1)_{ij} = \sum_{i=0}^{p_1} \sum_{j=0}^{p_1} b_i b_j \lim_{n\to\infty} \frac{1}{n_i n_j} \operatorname{tr} \underset{\sim}{T}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{T}_0^{-1} \underset{\sim}{G}_j$$

$$= \lim_{n\to\infty} \operatorname{tr} \underset{\sim}{T}_0^{-1} \left(\sum_{i=0}^{p_1} \frac{b_i}{n_i} \underset{\sim}{G}_i \right) \underset{\sim}{T}_0^{-1} \left(\sum_{j=0}^{p_1} \frac{b_j}{n_j} \underset{\sim}{G}_j \right)$$

because finite sums and limits and tracing interchange,

$$= \lim_{n\to\infty} \operatorname{tr}\left[\underset{\sim}{T}_0^{-1}\left(\sum_{i=0}^{p_1} \frac{b_i}{n_i} \underset{\sim}{G}_i\right)\right]^2$$

$$= \lim_{n\to\infty} \sum_{k=1}^{n} \lambda_k^2 \left[\underset{\sim}{T}_0^{-1}\left(\sum_{i=0}^{p_1} \frac{b_i}{n_i} \underset{\sim}{G}_i\right)\right] .$$

Thus the object of attention is positive for any finite n and the only problem is that it might degenerate in the limit. The proof that the limit is indeed positive proceeds as follows.

Suppose $b_0 \neq 0$; then some of the characteristic roots of

$\underset{\sim}{T}_0^{-1}\left(\sum_{i=0}^{p_1} \frac{b_i}{n_i} \underset{\sim}{G}_i\right)$ can be identified. Without loss of generality, write

$\underset{\sim}{T}_0 = \underset{\sim}{\Sigma}_0 = \sum_{i=0}^{p_1} \sigma_{0i} \underset{\sim}{G}_i$ for the rest of this proof. To find some of the characteristic vectors in this case note that there is a space of dimension $\nu_0(=n_0^2)$ orthogonal to $\mathcal{L}(\underset{\sim}{U}_1:\ldots:\underset{\sim}{U}_{p_1})$. (See Section 4.2.) Let an orthonormal basis for this space be $\underset{\sim}{H}_0$. Then $\underset{\sim}{U}_i'\underset{\sim}{H}_0 = \underset{\sim}{0}$ and hence $\underset{\sim}{G}_i\underset{\sim}{H}_0 = \underset{\sim}{0}$ for $i=1,2,\ldots,p_1$. This yields

$$\left(\sum_{i=0}^{p_1} \frac{b_i}{n_i} \underset{\sim}{G}_i\right)\underset{\sim}{H}_0 = \frac{b_0}{n_0} \underset{\sim}{I} \cdot \underset{\sim}{H}_0$$

since $G_0 = I$,

$$= \frac{b_0}{\sigma_{00} n_0} (\sigma_{00} \underset{\sim}{I})\underset{\sim}{H}_0$$

$$= \frac{b_0}{\sigma_{00} n_0} \left(\sum_{i=0}^{p_1} \sigma_{0i} \underset{\sim}{G}_i\right)\underset{\sim}{H}_0$$

$$= \frac{b_0}{\sigma_{00} n_0} \underset{\sim}{\Sigma}_0 \underset{\sim}{H}_0 .$$

Thus $\underset{\sim}{\Sigma}_0^{-1}\left(\sum_{i=0}^{p_1} \frac{b_i}{n_i} \underset{\sim}{G}_i\right)\underset{\sim}{H}_0 = \frac{b_0}{\sigma_{00} n_0} \underset{\sim}{H}_0$; then there are $\nu_0 (=n_0^2)$ characteristic

vectors whose characteristic roots are $\frac{b_0}{\sigma_{00} n_0}$. Then

$$\operatorname{tr}\left[\underset{\sim}{\Sigma}_0^{-1}\left(\sum_{i=0}^{p_1} \frac{b_i}{n_i} \underset{\sim}{G}_i\right)\right]^2 \geq n_0^2 \left[\frac{b_0}{\sigma_{00} n_0} \right]^2$$

$$= \frac{b_0^2}{\sigma_{00}^2} > 0$$

Now let $b_j = 0$ for all $j \notin S_s^*$; i.e. let s be the smallest index such that $b_i \neq 0$ for some $i \in S_s$. Now observe $\frac{|b_i|}{n_i}$ for $i \in S_s$. One

of these must be the largest in the sense that $\lim_{n\to\infty} \frac{|b_j|}{n_j} \cdot \frac{n_i}{|b_i|} \leq 1$

for $j \in S_s$ $j \neq i$. Consider this i fixed and now consider vectors belonging to $\mathcal{L}(\underset{\sim}{U}_i)$. Also without loss of generality let $b_i > 0$ (clearly $b_i \neq 0$). Now use Lemma B.5 to show that a number of characteristic roots of $\underset{\sim}{\Sigma}_0^{-1}\left(\sum_{j=1}^{p_1} \frac{b_j}{n_j} \underset{\sim}{G}_j\right)$ have a lower bound of the proper order of magnitude; that is, consider

$$\inf_{\substack{\underset{\sim}{x} \neq \underset{\sim}{0} \\ \underset{\sim}{x} \in L}} \frac{\underset{\sim}{x}'\left(\sum_{j=0}^{p_1} \frac{b_j}{n_j} \underset{\sim}{G}_j\right)\underset{\sim}{x}}{\underset{\sim}{x}'\underset{\sim}{\Sigma}_0\underset{\sim}{x}}$$

where L is a subspace of $\mathcal{L}(\underset{\sim}{U}_i)$. Equivalently, consider

$$\inf_{\underset{\sim}{y} \neq \underset{\sim}{0}} \frac{\underset{\sim}{y}'\underset{\sim}{U}_i'\left(\sum_{j=0}^{p_1} \frac{b_j}{n_j} \underset{\sim}{G}_j\right)\underset{\sim}{U}_i\underset{\sim}{y}}{\underset{\sim}{y}'\underset{\sim}{U}_i'\left(\sum_{j=0}^{p_1} \sigma_{0j}\underset{\sim}{G}_j\right)\underset{\sim}{U}_i\underset{\sim}{y}} ,$$

where restrictions are placed on possible $\underset{\sim}{y}$ vectors to restrict consideration to the appropriate subspace. $\mathcal{L}(\underset{\sim}{U}_i)$ is a space of dimension m_i; the number of restrictions placed on $\underset{\sim}{y}$ will determine how many characteristic roots there are greater than the lower bound which is eventually arrived at.

First, note that for any $j \in S^*_{s+1}$ (recall $i \in S_s$) the matrix $\underset{\sim}{U}'_j \underset{\sim}{U}_i$ is an $m_j \times m_i$ matrix and hence has rank at most m_j (because m_j has smaller order of magnitude than m_i). Hence by restricting $\underset{\sim}{\gamma}$ to an $m_i - m_j$ dimensional space it can be insured that $\underset{\sim}{U}'_j \underset{\sim}{U}_i \underset{\sim}{\gamma} = \underset{\sim}{0}$. Thus by restricting $\underset{\sim}{\gamma}$ to a space of at worst (i.e. smallest) dimension $m_i - \sum_{j \in S^*_{s+1}} m_j$ it can be insured that $\underset{\sim}{U}'_j \underset{\sim}{U}_i \underset{\sim}{\gamma} = 0$ and hence that $\underset{\sim}{G}_j \underset{\sim}{U}_i \underset{\sim}{\gamma} = \underset{\sim}{0}$ for $j \in S^*_{s+1}$. This restricts consideration to

$$\inf_{\substack{\underset{\sim}{\gamma} \neq \underset{\sim}{0} \\ \underset{\sim}{\gamma} \text{ restricted}}} \frac{\underset{\sim}{\gamma}' \underset{\sim}{U}'_i \left(\sum_{j \in S_s} \frac{b_j}{n_j} \underset{\sim}{G}_j \right) \underset{\sim}{U}_i \underset{\sim}{\gamma}}{\underset{\sim}{\gamma}' \underset{\sim}{U}'_i \left(\sum_{j \notin S^*_{s+1}} \sigma_{0j} \underset{\sim}{G}_j \right) \underset{\sim}{U}_i \underset{\sim}{\gamma}} ,$$

because $b_j = 0$ for $j \notin S^*_s$. Now note that $\underset{\sim}{\gamma}' \underset{\sim}{U}'_i \underset{\sim}{G}_i \underset{\sim}{U}_i \underset{\sim}{\gamma} = \underset{\sim}{\gamma}' \underset{\sim}{U}'_i \underset{\sim}{U}_i \underset{\sim}{U}'_i \underset{\sim}{U}_i \underset{\sim}{\gamma} = \underset{\sim}{\gamma}' \underset{\sim}{D}^2_i \underset{\sim}{\gamma} > 0$ because $\underset{\sim}{D}_i$ is a nonsingular diagonal matrix by the definition of $\underset{\sim}{U}_i$. Thus the expression under consideration can be rewritten as follows:

$$\inf_{\substack{\underset{\sim}{\gamma} \neq 0 \\ \underset{\sim}{\gamma} \text{ restricted}}} \frac{\frac{b_i}{n_i} \underset{\sim}{\gamma}' \underset{\sim}{U}'_i \underset{\sim}{G}_i \underset{\sim}{U}_i \underset{\sim}{\gamma} + \sum_{\substack{j \in S_s \\ j \neq i}} \frac{b_j}{n_j} \underset{\sim}{\gamma}' \underset{\sim}{U}'_i \underset{\sim}{G}_j \underset{\sim}{U}_i \underset{\sim}{\gamma}}{\sum_{\substack{j \notin S^*_{s+1} \\ j \neq i \\ j \neq 0}} \sigma_{0j} \underset{\sim}{\gamma}' \underset{\sim}{U}'_i \underset{\sim}{G}_j \underset{\sim}{U}_i \underset{\sim}{\gamma} + \sigma_{00} \underset{\sim}{\gamma}' \underset{\sim}{U}'_i \underset{\sim}{U}_i \underset{\sim}{\gamma} + \sigma_{0i} \underset{\sim}{\gamma}' \underset{\sim}{U}'_i \underset{\sim}{G}_i \underset{\sim}{U}_i \underset{\sim}{\gamma}}$$

$$= \inf_{\substack{\underset{\sim}{\gamma} \neq \underset{\sim}{0} \\ \underset{\sim}{\gamma} \text{ restricted}}} \frac{b_i}{n_i} \cdot \frac{1 + \sum\limits_{\substack{j \epsilon S_s \\ j \neq i}} \frac{b_j}{n_j} \frac{n_i}{b_i} \frac{\underset{\sim}{\gamma}' \underset{\sim}{U}_i' \underset{\sim}{U}_j \underset{\sim}{U}_j' \underset{\sim}{U}_i \underset{\sim}{\gamma}}{\underset{\sim}{\gamma}' \underset{\sim}{D}_i^2 \underset{\sim}{\gamma}}}{\sum\limits_{\substack{j \notin S^*_{s+1} \\ j \neq i \\ j \neq 0}} \sigma_{0j} \frac{\underset{\sim}{\gamma}' \underset{\sim}{U}_i' \underset{\sim}{U}_j \underset{\sim}{U}_j' \underset{\sim}{U}_i \underset{\sim}{\gamma}}{\underset{\sim}{\gamma}' \underset{\sim}{D}_i^2 \underset{\sim}{\gamma}} + \sigma_{00} \frac{\underset{\sim}{\gamma}' \underset{\sim}{D}_i \underset{\sim}{\gamma}}{\underset{\sim}{\gamma}' \underset{\sim}{D}_i^2 \underset{\sim}{\gamma}} + \sigma_{0i}}$$

by dividing each term by $\underset{\sim}{\gamma}' \underset{\sim}{D}_i^2 \underset{\sim}{\gamma}$ and factoring out $\frac{b_i}{n_i}$,

$$= \inf_{\substack{\underset{\sim}{\gamma} \neq 0 \\ \underset{\sim}{\gamma} \text{ restricted}}} \frac{b_i}{n_i} \frac{1 + \sum\limits_{\substack{j \epsilon S_s \\ j \neq i}} \frac{b_j n_i}{n_j b_i} \frac{\underset{\sim}{\gamma}' \underset{\sim}{D}_i \underset{\sim}{D}_i^{-1} \underset{\sim}{U}_i' \underset{\sim}{U}_j \underset{\sim}{U}_j' \underset{\sim}{U}_i \underset{\sim}{D}_i^{-1} \underset{\sim}{D}_i \underset{\sim}{\gamma}}{\underset{\sim}{\gamma}' \underset{\sim}{D}_i^2 \underset{\sim}{\gamma}}}{\sum\limits_{\substack{j \notin S^*_{s+1} \\ j \neq i \\ j \neq 0}} \sigma_{0j} \frac{\underset{\sim}{\gamma}' \underset{\sim}{D}_i \underset{\sim}{D}_i^{-1} \underset{\sim}{U}_i' \underset{\sim}{U}_j \underset{\sim}{U}_j' \underset{\sim}{U}_i \underset{\sim}{D}_i^{-1} \underset{\sim}{D}_i \underset{\sim}{\gamma}}{\underset{\sim}{\gamma}' \underset{\sim}{D}_i^2 \underset{\sim}{\gamma}} + \sigma_{00} \frac{\underset{\sim}{\gamma}' \underset{\sim}{D}_i \underset{\sim}{D}_i^{-1} \underset{\sim}{D}_i \underset{\sim}{\gamma}}{\underset{\sim}{\gamma}' \underset{\sim}{D}_i^2 \underset{\sim}{\gamma}} + \sigma_{0i}}$$

$$= \inf_{\substack{\underset{\sim}{\xi} \neq 0 \\ \underset{\sim}{\xi} \text{ restricted}}} \frac{b_i}{n_i} \frac{1 + \sum\limits_{\substack{j \epsilon S_s \\ j \neq i}} \frac{b_j n_i}{n_j b_i} \frac{\underset{\sim}{\xi}' \underset{\sim}{D}_i^{-1} \underset{\sim}{U}_i' \underset{\sim}{U}_j \underset{\sim}{U}_j' \underset{\sim}{U}_i \underset{\sim}{D}_i^{-1} \underset{\sim}{\xi}}{\underset{\sim}{\xi}' \underset{\sim}{\xi}}}{\sum\limits_{\substack{j \notin S^*_{s+1} \\ j \neq i \\ j \neq 0}} \sigma_{0j} \frac{\underset{\sim}{\xi}' \underset{\sim}{D}_i^{-1} \underset{\sim}{U}_i' \underset{\sim}{U}_j \underset{\sim}{U}_j' \underset{\sim}{U}_i \underset{\sim}{D}_i^{-1} \underset{\sim}{\xi}}{\underset{\sim}{\xi}' \underset{\sim}{\xi}} + \sigma_{00} \frac{\underset{\sim}{\xi}' \underset{\sim}{D}_i^{-1} \underset{\sim}{\xi}}{\underset{\sim}{\xi}' \underset{\sim}{\xi}} + \sigma_{0i}}$$

where $\underset{\sim}{\xi} = \underset{\sim}{D}_i \underset{\sim}{\gamma}$,

$$\geq \inf_{\substack{\xi \neq 0 \\ \xi \text{ restricted}}} \frac{b_i}{n_i} \frac{1 - \sum\limits_{\substack{j \in S_s \\ j \neq i}} \frac{|b_j| n_i}{|b_i| n_j} \frac{\xi' D_i^{-1} U_i' U_j U_j' U_i D_i^{-1} \xi}{\xi' \xi}}{\sum\limits_{\substack{j \notin S^*_{s+1} \\ j \neq i \\ j \neq 0}} \sigma_{0j} \frac{\xi' D_i^{-1} U_i' U_j U_j' U_i D_i^{-1} \xi}{\xi' \xi} + \sigma_{00} + \sigma_{0i}} \qquad (**)$$

because $\lambda_{max}(D_i^{-1}) \leq 1$ since each diagonal element of D_i is greater than 1. Thus matrices of the form $D_i^{-1} U_i' U_j U_j' U_i D_i^{-1}$ must be studied. Consider the trace of such a matrix.

$$\operatorname{tr}(D_i^{-1} U_i' U_j U_j' U_i D_i^{-1}) = \operatorname{tr}[(U_j' U_i D_i^{-1})' (U_j' U_i D_i^{-1})]$$

$$= \sum_{k=1}^{m_i} \sum_{\ell=1}^{m_j} (U_j' U_i D_i^{-1})^2_{\ell k} ,$$

because $\operatorname{tr} A'A = \sum_{\ell} \sum_{k} a^2_{\ell k}$ for any matrix A. But now let the columns of U_j be $u_1^{(j)}, \ldots, u_{m_j}^{(j)}$; then since $D_i = U_i' U_i$,

$$[U_j' U_i D_i^{-1}]_{\ell k} = \frac{u_\ell^{(j)\prime} \; u_k^{(i)}}{u_k^{(i)\prime} \; u_k^{(i)}} .$$

Observe that $\dfrac{\underset{\sim}{u}_\ell^{(j)\prime}\ \underset{\sim}{u}_k^{(i)}}{\underset{\sim}{u}_k^{(i)\prime}\ \underset{\sim}{u}_k^{(i)}} \leq 1$ because all columns contain only zeros and ones and hence the numerator counts matches of ones in $\underset{\sim}{u}_\ell^{(j)}$ and $\underset{\sim}{u}_k^{(i)}$ and the denominator counts the ones in $\underset{\sim}{u}_k^{(i)}$. Since there can be no more matches than there are ones in $\underset{\sim}{u}_k^{(i)}$ the inequality is true. Furthermore

$$\sum_{\ell=1}^{m_j} \frac{\underset{\sim}{u}_\ell^{(j)\prime}\ \underset{\sim}{u}_k^{(i)}}{\underset{\sim}{u}_k^{(i)\prime}\ \underset{\sim}{u}_k^{(i)}} = 1$$

because $\sum_{\ell=1}^{m_j} \underset{\sim}{u}_\ell^{(j)}$ is a nX1 vector of ones by definition of the $\underset{\sim}{U}_j$.

Hence $\sum_{\ell=1}^{m_j} \left(\dfrac{\underset{\sim}{u}_\ell^{(j)\prime}\ \underset{\sim}{u}_k^{(i)}}{\underset{\sim}{u}_k^{(i)\prime}\ \underset{\sim}{u}_k^{(i)}} \right)^2$ is a sum of squares of items all of which are between zero and one and which add up to one. Then the sum of squares has a maximum of one which occurs when one of the summands equals one and the others are zero. Otherwise the sum will be less than one (usually much less). At this point Assumption 4.2.4 quite naturally applies. It says that for $j \in S_s$, $j \neq i$ there exist constants R_1 and R_2 such that except for $R_1 m_i$ of the $\underset{\sim}{u}_k^{(i)}$, the quantities

$\sum_{\ell=1}^{m_j} \left(\dfrac{\underset{\sim}{u}_\ell^{(j)\prime}\ \underset{\sim}{u}_k^{(i)}}{\underset{\sim}{u}_k^{(i)\prime}\ \underset{\sim}{u}_k^{(i)}} \right)^2$ are less than R_2 . Thus

$$\mathrm{tr}(\underset{\sim}{D}_i^{-1}\underset{\sim}{U}_i'\underset{\sim}{U}_j\underset{\sim}{U}_j'\underset{\sim}{U}_i\underset{\sim}{D}_i^{-1}) = \sum_{k=1}^{m_i}\left[\sum_{\ell=1}^{m_j}\left(\frac{\underset{\sim}{u}_\ell^{(j)\prime}\;\underset{\sim}{u}_k^{(i)}}{\underset{\sim}{u}_k^{(i)\prime}\;\underset{\sim}{u}_k^{(i)}}\right)^2\right]$$

$$\leq (R_1 m_i)\,1 + (m_i - R_1 m_i)R_2$$

$$= m_i[R_1 + (1 - R_1)R_2]$$

$$\leq \frac{m_i}{N(S_s)+1}\;,$$

where $N(S_s)$ is the number of indices in S_s by Assumption 4.2.4. Of course, for $j \notin S_s$ the bound $\mathrm{tr}(\underset{\sim}{D}_i^{-1}\underset{\sim}{U}_i'\underset{\sim}{U}_j\underset{\sim}{U}_j'\underset{\sim}{U}_i\underset{\sim}{D}_i^{-1}) \leq m_i$ still holds,

since $\sum_{\ell=1}^{m_j}\left(\frac{\underset{\sim}{u}_\ell^{(j)\prime}\;\underset{\sim}{u}_k^{(i)}}{\underset{\sim}{u}_k^{(i)\prime}\;\underset{\sim}{u}_k^{(i)}}\right)^2 \leq 1,\ k=1,2,\ldots,m_i.$

Now if $\underset{\sim}{A}$ is any positive semidefinite m×m matrix with $\mathrm{tr}(\underset{\sim}{A}) \leq K_1 m$ it is clear that at most $\frac{K_1}{K_2}m$ of the characteristic roots of $\underset{\sim}{A}$ can be greater than K_2. (If they were, $\mathrm{tr}\,A > K_2 \cdot \frac{K_1}{K_2}m = K_1 m$, a contradiction.) This fact with $K_2 = 2p_1(N(S_s)+1)$ and $K_1 = 1$ implies that the denominator of (**) is less than $2p_1(N(S_s)+1)\sum_{\substack{j\in|S^*_{s+1}\\ j\neq i\\ j\neq 0}} \sigma_{0j} + \sigma_{00} + \sigma_{0i}$ if $\underset{\sim}{\xi}$

is restricted so that it contains no part of the $\frac{K_1}{K_2} m_i$ characteristic vectors that go along with the "offending" characteristic roots. Furthermore the numerator of (**) is equal to

$$1 - \sum_{\substack{j\epsilon S_s \\ j\neq i}} \left| \frac{b_j n_i}{b_i n_j} \right| \frac{\underset{\sim}{\xi}' \underset{\sim i}{D}^{-1} \underset{\sim i}{U}' \underset{\sim j}{U} \underset{\sim j}{U}' \underset{\sim i}{U} \underset{\sim i}{D}^{-1} \underset{\sim}{\xi}}{\underset{\sim}{\xi}' \underset{\sim}{\xi}}$$

$$\geq 1 - \left(1 + \frac{1}{2N(S_s)}\right) \sum_{\substack{j\epsilon S_s \\ j\neq i}} \frac{\underset{\sim}{\xi}' \underset{\sim i}{D}^{-1} \underset{\sim i}{U}' \underset{\sim j}{U} \underset{\sim j}{U}' \underset{\sim i}{U} \underset{\sim i}{D}^{-1} \underset{\sim}{\xi}}{\underset{\sim}{\xi}' \underset{\sim}{\xi}}$$

since $\left| \frac{b_j n_i}{b_i n_j} \right| \leq 1 + \frac{1}{2N(S_s)}$ for all $j\epsilon S_s$ $j\neq i$ for all n beyond some point as $n \to \infty$ because $\lim_{n\to\infty} \left| \frac{b_j n_i}{b_i n_j} \right| \leq 1$ by choice of i,

$$= 1 - \left(1 + \frac{1}{2N(S_s)}\right) \frac{\underset{\sim}{\xi}' [\underset{\sim i}{D}^{-1} \underset{\sim i}{U}' (\sum_{\substack{j\epsilon S_s \\ j\neq i}} \underset{\sim j}{U} \underset{\sim j}{U}') \underset{\sim i}{U} \underset{\sim i}{D}^{-1}] \underset{\sim}{\xi}}{\underset{\sim}{\xi}' \underset{\sim}{\xi}} .$$

But now

$$\mathrm{tr}[\underset{\sim i}{D}^{-1} \underset{\sim i}{U}' (\sum_{\substack{j\epsilon S_s \\ j\neq i}} \underset{\sim j}{U} \underset{\sim j}{U}') \underset{\sim i}{U} \underset{\sim i}{D}^{-1}] \leq \frac{N(S_s) - 1}{N(S_s) + 1} \cdot m_i$$

since the trace of a sum is the sum of the traces and since each individual trace is bounded above and there are $N(S_s)-1$ traces. Now using the above argument about traces and characteristic roots again with $K_2 = 1 - \frac{1}{2N(S_s)}$ and $K_1 = \frac{N(S_s) - 1}{N(S_s) + 1}$, if $\underset{\sim}{\xi}$ is further restricted so that it contains no part of the characteristic vectors associated with characteristic roots of $\underset{\sim}{D}_i^{-1}\underset{\sim}{U}_i'(\sum_{\substack{j\epsilon S_s \\ j\neq i}} \underset{\sim}{U}_j\underset{\sim}{U}_j')\underset{\sim}{U}_i\underset{\sim}{D}_i^{-1}$ which are greater than $1 - \frac{1}{2N(S_s)}$ then the numerator of (**) is greater than

$$1 - \left(1 + \frac{1}{2N(S_s)}\right)\left(1 - \frac{1}{2N(S_s)}\right) = \frac{1}{4N(S_s)^2} > 0.$$

It is now necessary to count the number of restrictions which have been made on $\underset{\sim}{\xi}$ at this point. At most $\sum_{j\epsilon S^*_{s+1}} m_j$ were made for the first restrictions (when $\underset{\sim}{\xi}$ was $\underset{\sim}{\gamma}$); at most

$$\sum_{\substack{j\notin S^*_{s+1} \\ j\neq i \\ j\neq 0}} \frac{K_1}{K_2} m_i = \sum_{\substack{j\notin S^*_{s+1} \\ j\neq i \\ j\neq 0}} \frac{m_i}{2p_1(N(S_s)+1)} \leq \frac{m_i}{2(N(S_s)+1)}$$ (because there are at most p_1 terms in the summation) more restrictions were made to bound the denominator; at most $\frac{K_1}{K_2} m_i = \frac{(N(S_s)-1)m_i}{(N(S_s)+1)(1-\frac{1}{2N(S_s)})}$ more restrictions were

made to bound the numerator. Thus the total number of restrictions is less than or equal to the sum of these numbers and the dimension of the linear space over which the inf may be taken will be greater than or equal to m_i minus this sum. In fact this dimension is greater than or equal to

$$m_i - \frac{m_i}{2(N(S_s)+1)} - \frac{m_i(N(S_s)-1)}{(N(S_s)+1)(1-\frac{1}{2N(S_s)})} - \sum_{j\epsilon S^*_{s+1}} m_j$$

$$= m_i\left[1 - \frac{1}{2(N(S_s)+1} - \frac{N(S_s)-1}{(N(S_s)+1)(1-\frac{1}{2N(S_s)})} - \sum_{j\epsilon S^*_{s+1}} \left(\frac{m_j}{m_i}\right)\right]$$

$$= m_i\left[\frac{2-\frac{1}{N(S_s)}}{2(N(S_s)+1)(1-\frac{1}{2N(S_s)})} - \sum_{j\epsilon S^*_{s+1}} \left(\frac{m_j}{m_i}\right)\right] .$$

But $\frac{m_j}{m_i} \to 0$ as $n \to \infty$ for each $j \epsilon S^*_{s+1}$ so that for n large enough the last expression in brackets is greater than $K_3 = \frac{1}{4}\left[\frac{2-\frac{1}{N(S_s)}}{(N(S_s)+1)(1-\frac{1}{2N(S_s)})}\right]$,

which is positive.

Now combine the bounds obtained on the numerator and denominator of (**) to obtain

$$\inf_{\substack{\underset{\sim}{\xi} \neq 0 \\ \underset{\sim}{\xi} \text{ restricted}}} \frac{b_i}{n_i} \frac{1 - \sum\limits_{\substack{j \in S_s \\ j \neq i}} \left| \frac{b_j n_i}{b_i n_j} \right| \frac{\underset{\sim}{\xi}' \underset{\sim}{D}_i^{-1} \underset{\sim}{U}_i' \underset{\sim}{U}_j \underset{\sim}{U}_j' \underset{\sim}{U}_i \underset{\sim}{D}_i^{-1} \underset{\sim}{\xi}}{\underset{\sim}{\xi}' \underset{\sim}{\xi}}}{\sum\limits_{\substack{j \notin S^*_{s+1} \\ j \neq i \\ j \neq 0}} \sigma_{0j} \frac{\underset{\sim}{\xi}' \underset{\sim}{D}_i^{-1} \underset{\sim}{U}_i' \underset{\sim}{U}_j \underset{\sim}{U}_j' \underset{\sim}{U}_i \underset{\sim}{D}_i^{-1} \underset{\sim}{\xi}}{\underset{\sim}{\xi}' \underset{\sim}{\xi}} + \sigma_{00} + \sigma_{0i}}$$

$$\geq \frac{b_i}{n_i} \frac{\frac{1}{4(N(S_s)^2)}}{2p_1(N(S_s)+1) \sum\limits_{\substack{j \in S^*_{s+1} \\ j \neq i \\ j \neq 0}} \sigma_{0j} + \sigma_{00} + \sigma_{0i}}$$

$$\equiv \frac{b_i}{n_i} K_4 > 0.$$

Now Lemma B.5 states that there are at least $K_3 m_i$ characteristic roots of $\underset{\sim}{\Sigma}_0^{-1}\left(\sum\limits_{j=0}^{p_1} \frac{b_j}{n_j} \underset{\sim}{G}_j\right)$ which are greater than or equal to $\frac{b_i}{n_i} K_4$.

This yields

$$\mathrm{tr}\left[\left(\underset{\sim}{\Sigma}_0^{-1}\left(\sum_{j=0}^{p_1} \frac{b_j}{n_j} \underset{\sim}{G}_j\right)\right]^2 \geq K_3 m_i \frac{b_i^2}{n_i^2} K_4^2$$

$$= \frac{m_i}{\nu_i} b_i^2 K_3 K_4^2$$

$$\underset{n\to\infty}{\to} \frac{1}{r_i} b_i^2 K_3 K_4^2 > 0 .$$

Thus for all choices of $\underset{\sim}{b}$, $\underset{\sim}{b}'\underset{\sim}{C}_1\underset{\sim}{b} > 0$ and hence $\underset{\sim}{C}_1$ is positive definite. This proves that $\underset{\sim}{J}$ is positive definite and concludes the proof of Lemma A.4.1. |||

A.4.2. Verification of Condition 3.3.1.ii--Asymptotic Normality of

$$\left.\frac{\partial\lambda}{\partial\underset{\sim}{\psi}}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}$$

LEMMA A.4.2. For $\underset{\sim}{\psi}$, $\lambda(\underset{\sim}{y},\underset{\sim}{\psi})$, and $\underset{\sim}{J}$ as defined in Section 4.5,

$$\left.\frac{\partial\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\underset{\sim}{\psi}}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} \xrightarrow{d} \underset{\sim}{\mathcal{N}}_p(\underset{\sim}{0},\underset{\sim}{J}) .$$

PROOF.

Let the vector $\left.\frac{\partial\lambda}{\partial\underset{\sim}{\psi}}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} = \underset{\sim}{a}_n = \begin{pmatrix} \underset{\sim}{a}_n^{(0)} \\ \underset{\sim}{a}_n^{(1)} \end{pmatrix}$, where $\underset{\sim}{a}_n^{(0)}$ is a p_0x1 vector defined by

$$\underset{\sim}{a}_n^{(0)} = \left.\frac{\partial\lambda}{\partial\underset{\sim}{\beta}}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}$$

$$= \frac{1}{n_{p_1+1}} \underset{\sim}{X}'\underset{\sim}{T}_0^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\,\frac{\underset{\sim}{\beta}_0}{n_{p_1+1}}),$$

and $\underset{\sim}{a}_n^{(1)}$ is a (p_1+1)x1 vector defined by

$$[\underset{\sim}{a}_n^{(1)}]_i = \left.\frac{\partial\lambda}{\partial\sigma_i}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}$$

$$= \frac{1}{2n_i}\left[-\text{tr}\ \underset{\sim}{T}_0^{-1}\underset{\sim}{G}_i + \left(\underset{\sim}{y} - \frac{\underset{\sim}{\beta}_0}{n_{p_1+1}}\right)'\underset{\sim}{T}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_0^{-1}\left(\underset{\sim}{y} - \frac{\underset{\sim}{\beta}_0}{n_{p_1+1}}\right)\right],$$

$$i=0,1,\ldots,p_1.$$

But

$$\underset{\sim}{T}_0 \equiv \sum_{i=0}^{p_1} \frac{[\underset{\sim}{\tau}_{0n}]_i}{n_i}\underset{\sim}{G}_i$$

$$= \sum_{i=0}^{p_1} \sigma_{0i}\underset{\sim}{G}_i$$

$$= \underset{\sim}{\Sigma}_0\ ,$$

and

$$\frac{\underset{\sim}{\beta}_0}{n_{p_1+1}} \equiv \frac{\underset{\sim}{\beta}_{0n}}{n_{p_1+1}}$$

$$= \underset{\sim}{\alpha}_0\ ,$$

for all n, so in the remainder of the proof, those substitutions will be used. Since $\underset{\sim}{\Sigma}_0$ is positive definite, let $\underset{\sim}{\Sigma}_0 = \underset{\sim}{A}\underset{\sim}{A}'$ which implies $\underset{\sim}{\Sigma}_0^{-1} = \underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}$. (Recall $\underset{\sim}{A}^{-t} \equiv (\underset{\sim}{A}')^{-1} = (\underset{\sim}{A}^{-1})'$.) Let $\underset{\sim}{z}$ be an nx1 vector defined by $\underset{\sim}{z} \equiv \underset{\sim}{A}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$; then $\underset{\sim}{z} \sim \mathcal{N}_n(\underset{\sim}{0},\underset{\sim}{I}_n)$. $\underset{\sim}{a}_n^{(0)}$ and $\underset{\sim}{a}_n^{(1)}$ are re-defined in terms of $\underset{\sim}{z}$ as follows:

$$\underset{\sim}{a}_n^{(0)} = \frac{1}{n_{p_1+1}} \underset{\sim}{X}'\underset{\sim}{A}^{-t}\underset{\sim}{z} \quad ,$$

$$[\underset{\sim}{a}_n^{(1)}]_i = \frac{1}{2n_i} (\underset{\sim}{z}'\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{z} - \mathrm{tr}\ \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i),\ i=0,1,\ldots,p_1.$$

Let $\delta_{p\times 1} = \begin{pmatrix} \underset{\sim}{\delta}^{(0)} \\ \underset{\sim}{\delta}^{(1)} \end{pmatrix}$, partitioned the same as $\underset{\sim}{a}_n$. It is then sufficient to show that for any choice of $\underset{\sim}{\delta}$, $\underset{\sim}{\delta}'\underset{\sim}{a}_n$ is asymptotically normal with mean zero and variance $\underset{\sim}{\delta}'\underset{\sim}{J}\underset{\sim}{\delta}$.

Let

$$W(\underset{\sim}{\delta},n) = \underset{\sim}{\delta}'\underset{\sim}{a}_n = \sum_{i=0}^{p_1} \frac{\delta_i^{(1)}}{2n_i} [\underset{\sim}{z}'(\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t})\underset{\sim}{z} - \mathrm{tr}\ \Sigma_0^{-1}\underset{\sim}{G}_i] + \frac{\underset{\sim}{\delta}^{(0)\prime}}{n_{p_1+1}} \underset{\sim}{X}'\underset{\sim}{A}^{-t}\underset{\sim}{z}$$

$$\equiv \underset{\sim}{z}'\underset{\sim}{F}(\underset{\sim}{\delta},n)\underset{\sim}{z} - \mathrm{tr}\ \underset{\sim}{F}(\underset{\sim}{\delta},n) + \underset{\sim}{f}'(\underset{\sim}{\delta},n)\underset{\sim}{z} \ ,$$

where $\quad \underset{\sim}{F}(\underset{\sim}{\delta},n) = \underset{\sim}{A}^{-1}\left(\sum_{i=0}^{p_1} \frac{\delta_i^{(1)}}{2n_i} \underset{\sim}{G}_i\right)\underset{\sim}{A}^{-t}$ and $\underset{\sim}{f}(\underset{\sim}{\delta},n) = \frac{1}{n_{p_1+1}} \underset{\sim}{A}^{-1}\underset{\sim}{X}\underset{\sim}{\delta}^{(0)}$.

Now calculate the characteristic function of $W(\underset{\sim}{\delta},n)$. For each n there exists an orthogonal matrix $\underset{\sim}{P}(\underset{\sim}{\delta},n)$ and a diagonal matrix $\underset{\sim}{\Lambda}(\underset{\sim}{\delta},n)$ ($[\underset{\sim}{\Lambda}(\underset{\sim}{\delta},n)]_{kk} = \lambda_k$) such that $\underset{\sim}{F}(\underset{\sim}{\delta},n) = \underset{\sim}{P}'(\underset{\sim}{\delta},n)\underset{\sim}{\Lambda}(\underset{\sim}{\delta},n)\underset{\sim}{P}(\underset{\sim}{\delta},n)$. Of course, $\underset{\sim}{\Lambda}$ contains the characteristic roots of $\underset{\sim}{F}$ and $\underset{\sim}{P}$ the characteristic vectors; the decomposition is possible because $\underset{\sim}{F}$ is symmetric.

Then the nx1 vector $\underset{\sim}{w}(\underset{\sim}{\delta},n) \equiv \underset{\sim}{P}(\underset{\sim}{\delta},n)\underset{\sim}{z} \sim \mathcal{N}_n(\underset{\sim}{0},\underset{\sim}{I}_n)$ and $\underset{\sim}{W}(\underset{\sim}{\delta},n) = \underset{\sim}{w}'(\underset{\sim}{\delta},n)\underset{\sim}{\Lambda}(\underset{\sim}{\delta},n)\underset{\sim}{w}(\underset{\sim}{\delta},n)-\text{tr }\underset{\sim}{\Lambda}(\underset{\sim}{\delta},n) + \underset{\sim}{g}'(\underset{\sim}{\delta},n)\underset{\sim}{w}(\underset{\sim}{\delta},n)$, where $\underset{\sim}{g}(\underset{\sim}{\delta},n) \equiv \underset{\sim}{P}(\underset{\sim}{\delta},n)\underset{\sim}{f}(\underset{\sim}{\delta},n)$ and $\text{tr }\underset{\sim}{F}(\underset{\sim}{\delta},n) = \text{tr }\underset{\sim}{\Lambda}(\underset{\sim}{\delta},n)$. Now the dependence on $\underset{\sim}{\delta}$ and n is suppressed in the notation, yielding $W = \underset{\sim}{w}'\underset{\sim}{\Lambda}\underset{\sim}{w} - \text{tr}\underset{\sim}{\Lambda} + \underset{\sim}{g}'\underset{\sim}{w}$ where $\underset{\sim}{w} \sim \mathcal{N}_n(\underset{\sim}{0},\underset{\sim}{I}_n)$. The characteristic function of W, $\phi_W(t)$ is given, for any t, by

$$\phi_W(t) = \mathcal{E}_0\{e^{itW}\}$$

$$= \mathcal{E}_0\{e^{it(\underset{\sim}{w}'\underset{\sim}{\Lambda}\underset{\sim}{w} - \text{tr }\underset{\sim}{\Lambda} + \underset{\sim}{g}'\underset{\sim}{w})}\}$$

$$= e^{-i\text{ tr }\underset{\sim}{\Lambda}}\,\mathcal{E}_0\{e^{i(t\underset{\sim}{g}'\underset{\sim}{w} + t\underset{\sim}{w}'\underset{\sim}{\Lambda}\underset{\sim}{w})}\}$$

Thus Lemma B.2 applies, yielding

$$\phi_W(t) = e^{-i\text{ tr }\underset{\sim}{\Lambda}}|\underset{\sim}{I}-2it\,\underset{\sim}{\Lambda}|^{-\frac{1}{2}}\, e^{-\frac{1}{2}t^2\underset{\sim}{g}'(\underset{\sim}{I}-2it\,\underset{\sim}{\Lambda})^{-1}\underset{\sim}{g}} .$$

Then $\log \phi_W(t) = -it\text{ tr }\underset{\sim}{\Lambda}-\frac{1}{2}\log|\underset{\sim}{I}-2it\,\underset{\sim}{\Lambda}|-\frac{1}{2}t^2\underset{\sim}{g}'(\underset{\sim}{I}-2it\,\underset{\sim}{\Lambda})^{-1}\underset{\sim}{g}$

$$= -it\sum_{k=1}^{n}\lambda_k - \frac{1}{2}\sum_{k=1}^{n}\log(1-2it\,\lambda_k)-\frac{1}{2}t^2\underset{\sim}{g}'\left(\sum_{j=0}^{\infty}(2it\,\underset{\sim}{\Lambda})^j\right)\underset{\sim}{g},$$

where the last expansion is valid so long as $\max_{k=1,2,\ldots,n}|\lambda_k| < \frac{1}{2t}$, which is true for n sufficiently large. Note that

$$\lambda_k = \lambda_k(\underset{\sim}{F}) = \lambda_k\left(\underset{\sim}{A}^{-t} \sum_{i=0}^{p_1} \frac{\delta_i^{(1)}}{2n_i} \underset{\sim}{G}_i \underset{\sim}{A}^{-1}\right)$$

$$= \lambda_k\left(\underset{\sim}{A}^{-1} \underset{\sim}{A}^{-t} \sum_{i=0}^{p_1} \frac{\delta_i^{(1)}}{2n_i} \underset{\sim}{G}_i\right)$$

by Lemma B.9,

$$= \lambda_k\left(\underset{\sim}{\Sigma}_0^{-1} \sum_{i=0}^{p_1} \frac{\delta_i^{(1)}}{2n_i} \underset{\sim}{G}_i\right)$$

$$= \frac{\sum_{i=0}^{p_1} \frac{\delta_i^{(1)}}{2n_i} \underset{\sim}{x}_k' \underset{\sim}{G}_i \underset{\sim}{x}_k}{\sum_{i=0}^{p_1} \sigma_{0i} \underset{\sim}{x}_k' \underset{\sim}{G}_i \underset{\sim}{x}_k}$$

by Lemma B.6, where $\underset{\sim}{x}_k$ is the associated characteristic vector.

Therefore

$$|\lambda_k| \leq \frac{\sum_{i=0}^{p_1} \left|\frac{\delta_i^{(1)}}{2n_i}\right| \underset{\sim}{x}_k' \underset{\sim}{G}_i \underset{\sim}{x}_k}{\sum_{i=0}^{p_1} \sigma_{0i} \underset{\sim}{x}_k' \underset{\sim}{G}_i \underset{\sim}{x}_k}$$

$$\leq \max_{i=0,1,\ldots,p_1} \frac{|\delta_i^{(1)}|}{2n_i \sigma_{0i}}$$

by Lemma B.3. The last quantity converges to zero since $n_i \to \infty$ for all i; this certainly can be made less than $\frac{1}{2t}$ for n large enough.

Continuing the expansion of $\log \phi_W(t)$,

$$\log \phi_W(t) = -it \sum_{k=1}^{n} \lambda_k - \tfrac{1}{2} \sum_{k=1}^{n} \log(1-2it\,\lambda_k) - \tfrac{1}{2} t^2 \underset{\sim}{g}' \sum_{j=0}^{\infty} (2it\,\underset{\sim}{\Lambda})^j \underset{\sim}{g}.$$

The second term is

$$-\tfrac{1}{2} \sum_{k=1}^{n} \log(1-2it\,\lambda_k) = \tfrac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{\infty} (2it\,\lambda_k)^j \cdot \frac{1}{j}$$

$$= \tfrac{1}{2} \sum_{k=1}^{n} [2it\,\lambda_k - 2t^2 \lambda_k^2 + \sum_{j=3}^{\infty} (2it\,\lambda_k)^j \cdot \frac{1}{j}],$$

while the last term is

$$-\tfrac{1}{2} t^2 \underset{\sim}{g}' \sum_{j=0}^{\infty} (2it\,\underset{\sim}{\Lambda}) \underset{\sim}{g} = -\tfrac{1}{2} t^2 \underset{\sim}{g}' \underset{\sim}{g} - \tfrac{1}{2} t^2 \sum_{j=1}^{\infty} \underset{\sim}{g}'(2it\,\underset{\sim}{\Lambda}) \underset{\sim}{g}.$$

Combining terms,

$$\log \phi_W(t) = \cancel{-it \sum_{k=1}^{n}} \lambda_k + \cancel{it \sum_{k=1}^{n}} \lambda_k - \frac{t^2}{2} \sum_{k=1}^{n} 2\lambda_k^2 - \frac{t^2}{2} \underset{\sim}{g}' \underset{\sim}{g}$$

$$+ \tfrac{1}{2} \sum_{k=1}^{n} \sum_{j=3}^{\infty} (2it\,\lambda_k)^j \frac{1}{j}$$

$$-\tfrac{1}{2} t^2 \sum_{j=1}^{\infty} (2it)^j \underset{\sim}{g}' \underset{\sim}{\Lambda}^j \underset{\sim}{g}.$$

But recall that

$$\underset{\sim}{g}'\underset{\sim}{g} = \underset{\sim}{f}'\underset{\sim}{P}'\underset{\sim}{P}\underset{\sim}{f}$$

$$= \underset{\sim}{f}'\underset{\sim}{f}$$

$$= \frac{1}{n^2_{p_1+1}} \; \underset{\sim}{\delta}^{(0)\prime}\underset{\sim}{X}'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{X}\underset{\sim}{\delta}^{(0)}$$

$$= \frac{1}{\nu_{p_1+1}} \underset{\sim}{\delta}^{(0)\prime}\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{X} \underset{\sim}{\delta}^{(0)}$$

$$\to \underset{\sim}{\delta}^{(0)\prime} \; \underset{\sim}{C}_0 \; \underset{\sim}{\delta}^{(0)}$$

by Assumption 4.2.5. Furthermore

$$2 \sum_{k=1}^{n} \lambda_k^2 = 2\mathrm{tr}\, \underset{\sim}{\Lambda}^2 = 2\mathrm{tr}\, \underset{\sim}{F}^2$$

$$= 2\mathrm{tr}\left(A^{-1} \sum_{i=0}^{p_1} \frac{\delta_i^{(1)}}{2n_i} G_i A^{-t} A^{-1} \sum_{j=0}^{p_1} \frac{\delta_j^{(1)}}{2n_j} G_j A^{-t}\right)$$

$$= \sum_{i=0}^{p_1} \sum_{j=0}^{p_1} \delta_i^{(1)} \delta_j^{(1)} \cdot \frac{1}{2n_i n_j} \mathrm{tr}\, \underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-1}\underset{\sim}{G}_j\underset{\sim}{A}^{-t}$$

$$= \sum_{i=0}^{p_1} \sum_{j=0}^{p_1} \delta_i^{(1)} \delta_j^{(1)} \frac{1}{2n_i n_j} \mathrm{tr}\, \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j$$

$$\rightarrow \sum_{i=0}^{p_1} \sum_{j=0}^{p_1} \delta_i^{(1)} \delta_j^{(1)} (\underset{\sim}{C}_1)_{ij}$$

$$= \underset{\sim}{\delta}^{(1)\prime} \underset{\sim}{C}_1 \underset{\sim}{\delta}^{(1)} .$$

Thus since $\underset{\sim}{J} = \begin{pmatrix} \underset{\sim}{C}_0 & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{C}_1 \end{pmatrix}$ the first remaining terms of $\log \phi_W(t)$ converge to $\frac{1}{2}t^2\underset{\sim}{\delta}'\underset{\sim}{J}\underset{\sim}{\delta}$ which is what is required for normality. It remains to show the last two terms converge to zero.

Now

$$S_1 \equiv \left| \sum_{k=1}^{n} \sum_{j=3}^{\infty} (2it\,\lambda_k)^j \frac{1}{j} \right| \leq \sum_{k=1}^{n} \sum_{j=3}^{\infty} 2^j t^j \, |\lambda_k|^j .$$

As previously noted $\max|\lambda_k| \rightarrow 0$ as $n \rightarrow \infty$ so that $2t \max|\lambda_k| < \frac{1}{2}$ for n large enough and therefore

$$S_1 \leq \sum_{k=1}^{n} \frac{2^3 t^3 |\lambda_k|^3}{1-2t|\lambda_k|}$$

$$\leq 16t^3 \sum_{k=1}^{n} |\lambda_k|^3$$

$$\leq 16t^3 \max |\lambda_k| \sum_{k=1}^{n} |\lambda_k|^2 .$$

But $\sum_{k=1}^{n} |\lambda_k|^2 \rightarrow \frac{1}{2} \underset{\sim}{\delta}^{(1)\prime} \underset{\sim}{C}_1 \underset{\sim}{\delta}^{(1)} < \infty$ as previously noted and is therefore

bounded, and $\max|\lambda_k| \to 0$ so that indeed $S_1 \to 0$.

$$S_2 \equiv \left| \sum_{j=1}^{\infty} (2it)^j \underset{\sim}{g}' \underset{\sim}{\Lambda}^j \underset{\sim}{g} \right|$$

$$= \left| \sum_{j=1}^{\infty} (2it)^j \underset{\sim}{f}' \underset{\sim}{P}' \underset{\sim}{\Lambda}^j \underset{\sim}{P} \underset{\sim}{f} \right|$$

$$= \left| \sum_{j=1}^{\infty} (2it)^j \underset{\sim}{f}' \underset{\sim}{F}^j \underset{\sim}{f} \right|$$

$$\leq \sum_{j=1}^{\infty} 2^j t^j \left| \underset{\sim}{f}' \underset{\sim}{F}^j \underset{\sim}{f} \right|$$

$$\leq \sum_{j=1}^{\infty} 2^j t^j \underset{\sim}{f}' \underset{\sim}{f} \lambda_{max}^{\frac{1}{2}} [(\underset{\sim}{F}')^j \underset{\sim}{F}^j]$$

by Lemma B.12. But $\underset{\sim}{f}'\underset{\sim}{f} \to \underset{\sim}{\delta}^{(0)}{}' \underset{\sim}{C}_0 \underset{\sim}{\delta}^{(0)} < \infty$ and is therefore bounded and $\underset{\sim}{F}$ is symmetric so that

$$\lambda_{max}[(\underset{\sim}{F}')^j \underset{\sim}{F}^j] = \lambda_{max}(\underset{\sim}{F}^{2j})$$

$$\leq \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{F})|^{2j}$$

by Lemma B.14, so that

$$\lambda_{max}^{\frac{1}{2}}[(\underset{\sim}{F}')^j \underset{\sim}{F}^j] \leq \max|\lambda_k(\underset{\sim}{F})|^j$$

$$= \max|\lambda_k|^j .$$

But again this converges to zero so that it may be assumed $2t \max|\lambda_k| < \frac{1}{2}$ and therefore

$$S_2 \le \underset{\sim}{f}'\underset{\sim}{f} \sum_{j=1}^{\infty} (2t \max|\lambda_k|)^j$$

$$= \underset{\sim}{f}'\underset{\sim}{f} \frac{2t \max|\lambda_k|}{1-2t \max|\lambda_k|}$$

$$\le 4t\, \underset{\sim}{f}'\underset{\sim}{f} \max|\lambda_k|$$

$$\to 0.$$

Thus $\log \phi_W(t) \to -\frac{1}{2}t^2\underset{\sim}{\delta}'\underset{\sim}{J}\underset{\sim}{\delta}$ and since $\underset{\sim}{J}$ is positive definite the limit is a legitimate characteristic function, continuous at $t=0$ and $\underset{\sim}{\delta}=\underset{\sim}{0}$. In fact $e^{-\frac{1}{2}t^2\underset{\sim}{\delta}'\underset{\sim}{J}\underset{\sim}{\delta}'}$ is the characteristic function of a random variable distributed as $\eta(0,\underset{\sim}{\delta}'\underset{\sim}{J}\underset{\sim}{\delta})$. This proves that $W(\underset{\sim}{\delta},n) \xrightarrow{d} \eta(0,\underset{\sim}{\delta}'\underset{\sim}{J}\underset{\sim}{\delta})$ and since $\underset{\sim}{\delta}$ is arbitrary that $\underset{\sim}{a}_n \xrightarrow{d} \underset{\sim}{\eta}_p(\underset{\sim}{0},\underset{\sim}{J})$, which was to be proved.|||

A.4.3. Verification of Condition 3.3.1.iii--Convergence in Probability of $\left.\frac{\partial^2\lambda}{\partial\underset{\sim}{\psi}\partial\underset{\sim}{\psi}'}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}$ to Its Expected Value

Lemma A.4.3. For $\underset{\sim}{\psi}$ and $\lambda(\underset{\sim}{y},\underset{\sim}{\psi})$ as defined in Section 4.5,

$\left.\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}$ converges in probability to its expected value,

$i,j=1,2,\ldots,p$.

PROOF.

It suffices to prove that the variance of each element of $\frac{\partial^2\lambda}{\partial\underset{\sim}{\psi}\partial\underset{\sim}{\psi}'}$ converges to zero as $n \to \infty$. There are three forms of second derivatives, which are exhibited in Section 4.3. Each will be dealt with separately.

Since each element of $\frac{\partial^2\lambda}{\partial\underset{\sim}{\beta}\partial\underset{\sim}{\beta}'}$ has variance zero (being a constant), the lemma is true for these derivatives.

To examine the derivatives of the form $\frac{\partial^2\lambda}{\partial\underset{\sim}{\beta}\partial\tau_i}$, it is sufficient to show that for all $p_0 \times 1$ vectors $\underset{\sim}{\xi}$ such that $\underset{\sim}{\xi}'\underset{\sim}{\xi} = 1$, $\mathrm{Var}_0\{\phi_i(\underset{\sim}{\xi})\} \to 0$, $i=0,1,\ldots,p_1$, where

$$\phi_i(\underset{\sim}{\xi}) = \underset{\sim}{\xi}'\left[\frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\underset{\sim}{\beta}\partial\tau_i} \bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} \right]$$

$$= \frac{1}{n_i n_{p_1+1}} \underset{\sim}{\xi}'\underset{\sim}{X}'\underset{\sim}{T}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_0^{-1}\left(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_0}{n_{p_1+1}}\right)$$

from Section 4.3,

$$= \frac{1}{n_i n_{p_1+1}} \underset{\sim}{\xi}'\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0),$$

using the same substitutions as in Section A.4.2.

Then $\mathcal{E}_0\{\phi_i(\underset{\sim}{\xi})\} = 0$ and

$$\text{Var}_0\{\phi_i(\underset{\sim}{\xi})\} = \frac{1}{n_i^2 n_{p_1+1}^2} \underset{\sim}{\xi}'\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{\Sigma}_0\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X}\underset{\sim}{\xi}$$

$$= \frac{1}{n_i^2 n_{p_1+1}^2} \underset{\sim}{\xi}'\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X}\underset{\sim}{\xi}$$

$$= \frac{1}{n_i^2 n_{p_1+1}^2} \underset{\sim}{\xi}'\underset{\sim}{X}'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{X}\underset{\sim}{\xi}$$

(Recall that $\underset{\sim}{\Sigma}_0 = \underset{\sim}{A}\underset{\sim}{A}'$.)

$$\le \frac{1}{n_{p_1+1}^2}\left[\underset{\sim}{\xi}'\underset{\sim}{X}'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{X}\underset{\sim}{\xi}\right] \frac{1}{n_i^2}\, \lambda_{\max}(\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{G}_i\underset{\sim}{A}^{-t})$$

by definition of characteristic root,

$$= \frac{1}{n_{p_1+1}^2} \left[\underset{\sim}{\xi}' \underset{\sim}{X}' \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{X} \underset{\sim}{\xi} \right] \frac{1}{n_i^2} \lambda_{max} (\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i)^2$$

by definition of $\underset{\sim}{A}$ and by Lemma B.9,

$$\leq \frac{\underset{\sim}{\xi}' \underset{\sim}{\xi}}{n_i^2} \left[\frac{1}{n_{p_1+1}^2} \lambda_{max} (\underset{\sim}{X}' \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{X}) \right] \lambda_{max} (\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i)^2$$

again by definition of characteristic root,

$$\leq \frac{B}{n_i^2} \cdot \frac{1}{\sigma_{0i}^2}$$

by Propositions A.3.2 and A.3.10. The last expression converges to zero as $n \to \infty$ because $n_i^2 = \nu_i$ and ν_i converges to infinity by Assumptions 4.2.1 and 4.2.3.

It remains to show that $\mathrm{Var}_0 \left\{ \left. \frac{\partial^2 \lambda(\underset{\sim}{y}, \underset{\sim}{\psi})}{\partial \tau_i \partial \tau_j} \right|_{\underset{\sim}{\psi} = \underset{\sim}{\psi}_{0n}} \right\} \to 0$ as $n \to \infty$, $i,j=0,1,\ldots,p_1$. But

$$\mathrm{Var}_0 \left\{ \left. \frac{\partial^2 \lambda(\underset{\sim}{y}, \underset{\sim}{\psi})}{\partial \tau_i \partial \tau_j} \right|_{\underset{\sim}{\psi} = \underset{\sim}{\psi}_{0n}} \right\} = \mathrm{Var}_0 \left\{ \frac{1}{2 n_i n_j} \left[\mathrm{tr}\ \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j \right. \right.$$

$$\left. \left. -2 (\underset{\sim}{y} - \underset{\sim}{X} \underset{\sim}{\alpha}_0)' \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j \underset{\sim}{\Sigma}_0^{-1} (\underset{\sim}{y} - \underset{\sim}{X} \underset{\sim}{\alpha}_0) \right] \right\}$$

using the same substitutions as above,

$$= \frac{2}{n_i^2 n_j^2} \operatorname{tr}(\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{\Sigma}_0)^2$$

by Lemma B.1,

$$\leq \frac{2(\min[m_i, m_j])}{n_i^2 n_j^2} \lambda_{\max}^2(\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j)$$

because there are at most $\min[m_i, m_j]$ nonzero characteristic roots of

$\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j$,

$$= \frac{2(\min[m_i, m_j])}{n_i^2 n_j^2} \lambda_{\max}^2(\underset{\sim}{A}^{-1} \underset{\sim}{G}_i \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{G}_j \underset{\sim}{A}^{-t})$$

by Lemma B.9,

$$\leq \frac{2(\min[m_i, m_j])}{n_i^2 n_j^2} \lambda_{\max}^2(\underset{\sim}{A}^{-1} \underset{\sim}{G}_i \underset{\sim}{A}^{-t}) \lambda_{\max}^2(\underset{\sim}{A}^{-1} \underset{\sim}{G}_j \underset{\sim}{A}^{-t})$$

by Lemma B.13,

$$\leq \frac{2}{\sigma_{0i}^2 \sigma_{0j}^2} \cdot \frac{\min[m_i, m_j]}{n_i^2 n_j^2}$$

by Lemma B.9 and Proposition A.3.2,

$$\to 0$$

by Proposition A.3.10.

Thus in all cases, $\operatorname{Var}_0 \left\{ \left. \frac{\partial^2 \lambda}{\partial \psi_i \partial \psi_j} \right|_{\psi = \psi_{0n}} \right\} \to 0$ as $n \to \infty$, which was to be proved. |||

A.4.4. Verification of Conditions 3.3.1.vi and 3.3.1.iv--Convergence in Probability to Zero of $\frac{\partial^2\lambda}{\partial\underset{\sim}{\psi}\partial\underset{\sim}{\psi}'}\Big|_{\underset{\sim}{\psi}_n} - \frac{\partial^2\lambda}{\partial\underset{\sim}{\psi}\partial\underset{\sim}{\psi}'}\Big|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}}$ Uniformly for $\underset{\sim}{\psi}_n \in S_b(\underset{\sim}{\psi}_{0n})$

LEMMA A.4.4. For $\underset{\sim}{\psi}$ and $\lambda(\underset{\sim}{y},\underset{\sim}{\psi})$ as defined in Section 4.5, and for any $b > 0$, if Conditions A.2.1 and A.3.1 are true, then

$$\sup_{\underset{\sim}{\psi}_n \in S_b(\underset{\sim}{\psi}_{0n})} \left| \frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\Big|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_n} - \frac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}\Big|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{0n}} \right| \to 0$$

as $n \to \infty$, $i,j=1,2,\ldots,p$.

PROOF.

Let $\underset{\sim}{\psi}_{1n}$ be any point in $S_b(\underset{\sim}{\psi}_{0n})$. Then for the first set of derivatives,

$$\frac{\partial^2\lambda}{\partial\underset{\sim}{\beta}\partial\underset{\sim}{\beta}'}\Big|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{an}} = -\frac{1}{n^2_{p_1+1}}\underset{\sim}{X}'\underset{\sim}{T}_a^{-1}\underset{\sim}{X}, \quad a=0,1.$$

(Recall $\underset{\sim}{T}_0 = \underset{\sim}{\Sigma}_0$.)

It is clearly sufficient for these derivatives to show for any $\underset{\sim}{\xi}_1, \underset{\sim}{\xi}_2$, $p_0 \times 1$ such that $\underset{\sim}{\xi}_1'\underset{\sim}{\xi}_1 = \underset{\sim}{\xi}_2'\underset{\sim}{\xi}_2 = 1$ that $\phi \equiv \frac{1}{n^2_{p_1+1}} |\underset{\sim}{\xi}_1'\underset{\sim}{X}'(\underset{\sim}{T}_1^{-1} - \underset{\sim}{\Sigma}_0^{-1})\underset{\sim}{X}\underset{\sim}{\xi}_2| \to 0$

independent of $\underset{\sim}{\psi}_{1n}$. But $\underset{\sim}{T}_1^{-1} - \underset{\sim}{T}_0^{-1} = \underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$ and $\underset{\sim}{\Sigma}_0^{-1} = \underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}$ as usual; then

$$\phi^2 = \left| \frac{1}{n_{p_1+1}^2} \underset{\sim}{\xi}_1' \underset{\sim}{X}'(\underset{\sim}{T}_1^{-1} - \underset{\sim}{\Sigma}_0^{-1})\underset{\sim}{X}\underset{\sim}{\xi}_2 \right|^2$$

$$\leq (\underset{\sim}{\xi}_1'\underset{\sim}{\xi})(\underset{\sim}{\xi}_2'\underset{\sim}{\xi}_2)\left[\frac{1}{n_{p_1+1}^2} \lambda_{max}(\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X})\right]^2 \lambda_{max}(\underset{\sim}{A}'\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{T}_1^{-1}\underset{\sim}{A})^2$$

by Proposition A.3.4 and the fact that $\underset{\sim}{T}_1^{-1} - \underset{\sim}{T}_0^{-1}$ is symmetric. But

$$\lambda_{max}(\underset{\sim}{A}'\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{T}_1^{-1}\underset{\sim}{A})^2 \leq \max_{k=1,2,\ldots,n} \left|\lambda_k[\underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)]\right|^2$$

by Lemmae B.11 and B.14,

$$\leq \frac{4b^2}{\min\limits_{i=0,1,\ldots,p_1} (n_i\sigma_{0i})^2}$$

by Proposition A.3.2. $\frac{1}{n_{p_1+1}^2} \lambda_{max}(\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X})$ is bounded by A.3.10 and thus $\phi \leq \dfrac{4b^2 B}{\min\limits_{i=0,1,\ldots,p_1}(n_i\sigma_{0i})^2}$ (B is taken from Proposition A.3.10) and hence $\phi \to 0$ independent of $\underset{\sim}{\psi}_{1n}$. (Since none of the bounds or convergence rates depend on $\underset{\sim}{\psi}_{1n}$.)

For the next set of derivatives,

$$\frac{\partial^2\lambda}{\partial\tau_i\partial\underset{\sim}{\beta}}\bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{an}} = \frac{1}{n_i n_{p_1+1}} \underset{\sim}{X}'\underset{\sim}{T}_a^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_a^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_a}{n_{p_1+1}}),$$

$$i=0,1,\ldots,p_1,\quad a=0,1.$$

Therefore, it is sufficient to show for these derivatives that for all $\underset{\sim}{\xi}$ $p_0\times 1$ such that $\underset{\sim}{\xi}'\underset{\sim}{\xi} = 1$ that

$$\left|\frac{1}{n_i n_{p_1+1}}\underset{\sim}{\xi}'\underset{\sim}{X}'\left[\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_1}{n_{p_1+1}})-\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_0}{n_{p_1+1}})\right]\right| \to 0$$

independent of $\underset{\sim}{\psi}_{1n}$. Lemma B.16 applies here and thus it is sufficient to show that each of the following two terms goes to zero independent of $\underset{\sim}{\psi}_{1n}$.

$$\phi_1 = \frac{1}{n_i n_{p_1+1}}\left|\underset{\sim}{\xi}'\underset{\sim}{X}'(\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1} - \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1})(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)\right|,$$

(Recall that $\frac{\underset{\sim}{\beta}_0}{n_{p_1+1}} = \underset{\sim}{\alpha}_0$.)

$$\phi_2 = \frac{1}{n_i n_{p_1+1}}\left|\underset{\sim}{\xi}'\underset{\sim}{X}'\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\underset{\sim}{X}\frac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)}{n_{p_1+1}}\right|.$$

Now

$$\phi_2^2 \le \frac{1}{n_i^2}(\underset{\sim}{\xi}'\underset{\sim}{\xi})(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)'(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)\left[\frac{1}{n_{p_1+1}^2}\lambda_{max}(\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X})\right]^2$$

$$\cdot\ \lambda_{max}(\underset{\sim}{A}'\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\underset{\sim}{A})$$

by Proposition A.3.4,

$$\leq \frac{1}{n_i^2}\cdot 1\cdot p_0 b^2 B^2\cdot\frac{1}{\sigma_{0i}^2}\ \lambda_{max}^4(\underset{\sim}{\Sigma}_0\underset{\sim}{T}_1^{-1})$$

by Propositions A.3.1, A.3.2, A.3.3, A.3.10,

$$\leq \frac{1}{n_i^2}\, p_0 b^2 B\, \frac{1}{\sigma_{0i}^2}\ 16$$

$$\to 0,\ i=0,1,\ldots,p_1,$$

independent of $\underset{\sim}{\psi}_{1n}$.

Now a technique is developed which is used often in this and subsequent sections. $\underset{\sim}{T}_1^{-1} = \underset{\sim}{\Sigma}_0^{-1} + \underset{\sim}{T}_1^{-1} - \underset{\sim}{\Sigma}_0^{-1} = \underset{\sim}{\Sigma}_0^{-1} + \underset{\sim}{\Delta}$, where $\underset{\sim}{\Delta} = \underset{\sim}{T}_1^{-1} - \underset{\sim}{\Sigma}_0^{-1}$. As noted previously, $\underset{\sim}{\Delta}$ is symmetric and $\underset{\sim}{\Delta} = \underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$; thus $\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1} - \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1} = \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Delta} + \underset{\sim}{\Delta}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1} + \underset{\sim}{\Delta}\underset{\sim}{G}_i\underset{\sim}{\Delta}$. ϕ_1 then breaks down into three terms, each of which fits into the following formula from Proposition A.3.4 with appropriate choice of $\underset{\sim}{F}_1$ and $\underset{\sim}{F}_2$.

$$\left[\frac{1}{n_i n_{p_1+1}}\,\underset{\sim}{\xi}'\underset{\sim}{X}'\underset{\sim}{F}_1\underset{\sim}{F}_2(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)\right]^2 \leq \frac{1}{n_i^2}\,\underset{\sim}{\xi}'\underset{\sim}{\xi}\left[\frac{1}{n_{p_1+1}^2}\lambda_{max}(\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X})\right]\lambda_{max}(\underset{\sim}{A}'\underset{\sim}{F}_1'\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A})$$

$$\cdot(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$$

$$\leq \frac{1}{n_i^2} B\, \lambda_{max}(\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A})(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$$

by Proposition A.3.10. Now fit in the $\frac{1}{n_i^2}$ where appropriate with $\underset{\sim}{F}_1$ and $\underset{\sim}{F}_2$ to obtain the division shown in Table A.4.4.1.

Table A.4.4.1

Division of ϕ_1 into Terms with Appropriate $\underset{\sim}{F}_1$ and $\underset{\sim}{F}_2$

Term	$\underset{\sim}{F}_1$	$\underset{\sim}{F}_2$
1	$\frac{1}{n_i}\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}$	$(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$
2	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$	$\frac{1}{n_i}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}$
3	$\frac{1}{n_i}\underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}$	$(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$

Proposition A.3.3 now yields the following bounds for $\lambda_{max}(\underset{\sim}{A}'\underset{\sim}{F}_1'\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1'\underset{\sim}{A})$ which are given in Table A.4.4.2.

Table A.4.4.2

Bounds for $\lambda_{max}(\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A})$ with $\underset{\sim}{F}_1$ from Table A.4.4.1

Term	Bound for $\lambda_{max}(\underset{\sim}{A}'\underset{\sim}{F}_1'\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A})$
1	$\dfrac{4}{n_i^2\sigma_{0i}^2}$
2	$\dfrac{4b^2}{\min\limits_{j=0,1,\ldots,p_1}(n_j\sigma_{0j})^2}$
3	$\dfrac{16b^2}{n_i^2\sigma_{0i}^2\min\limits_{j=0,1,\ldots,p_1}(n_j\sigma_{0j})^2}$

All these bounds go to zero independent of $\underset{\sim}{\psi}_{1n}$; thus it is sufficient to show that $(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$ is bounded independent of $\underset{\sim}{\psi}_{1n}$ for the two possible choices of $\underset{\sim}{F}_2$. The remarks preceeding Proposition A.3.5 show that it is sufficient to show that $\underset{\sim}{w}_s'\underset{\sim}{Q}_s'\underset{\sim}{A}'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2\underset{\sim}{A}\underset{\sim}{Q}_s\underset{\sim}{w}_s$ is bounded for $s=0,1,\ldots,c$. But propositions A.3.5 and A.3.6 apply here and together with Conditions A.2.1 and A.3.1 they yield the following bounds.

Table A.4.4.3

Bounds for $\underset{\sim}{w}_s'\underset{\sim}{Q}_s\underset{\sim}{A}'\underset{\sim}{F}_2\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2\underset{\sim}{A}\underset{\sim}{Q}_s\underset{\sim}{w}_s$ for $\underset{\sim}{F}_2$ from Table A.4.4.1

$\underset{\sim}{F}_2$	Bound for $\underset{\sim}{w}_s'\underset{\sim}{Q}_s\underset{\sim}{A}'\underset{\sim}{F}_2\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2\underset{\sim}{A}\underset{\sim}{Q}_s\underset{\sim}{w}_s$
$\frac{1}{n_i}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}$	$\begin{cases} 0 & i \in S_{s+1}^* \\ \frac{1}{\sigma_{0i}^2} \cdot \frac{11}{10} \frac{\tilde{m}_s}{n_i} & i \notin S_{s+1}^* \end{cases}$
$(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$	$b^2 \cdot \frac{11}{10} \cdot \frac{\tilde{m}_s}{\min\limits_{i=0,1,\ldots,i_{s+1}-1} (n_i\sigma_{0i})^2}$

But Proposition A.3.10 guarantees that the bounds in Table A.4.4.3 are themselves bounded and hence for each $i=0,1,\ldots,p_1$ the desired convergence to zero independent of $\underset{\sim}{\psi}_{1n}$ takes place.

It now remains to treat the terms $\frac{\partial^2\lambda}{\partial\tau_i\partial\tau_j}$.

$$\left.\frac{\partial^2\lambda}{\partial\tau_i\partial\tau_j}\right|_{\underset{\sim}{\psi}=\psi_{an}} = \frac{1}{2n_in_j}\left[\left(\text{tr}\ \underset{\sim}{T}_a^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_a^{-1}\underset{\sim}{G}_j - 2(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_a}{n_{p_1+1}})'\underset{\sim}{T}_a^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_a^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_a^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_a}{n_{p_1+1}}\right)\right],$$

$i,j=0,1,\ldots,p_1$, $a=0,1$.

It is sufficient for these derivatives to show that

$$\phi_0 = \frac{1}{2n_i n_j} \left| \operatorname{tr} \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_i \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_j - \operatorname{tr} \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j \right|$$

and

$$\phi_1 = \frac{1}{n_i n_j} \left| \left(\underset{\sim}{y} - \underset{\sim}{X} \frac{\underset{\sim}{\beta}_1}{n_{p_1+1}} \right)' \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_i \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_j \underset{\sim}{T}_1^{-1} \left(\underset{\sim}{y} - \underset{\sim}{X} \frac{\underset{\sim}{\beta}_1}{n_{p_1+1}} \right) \right.$$

$$\left. - \left(\underset{\sim}{y} - \underset{\sim}{X} \frac{\underset{\sim}{\beta}_0}{n_{p_1+1}} \right)' \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j \underset{\sim}{\Sigma}_0^{-1} \left(\underset{\sim}{y} - \underset{\sim}{X} \frac{\underset{\sim}{\beta}_0}{n_{p_1+1}} \right) \right|$$

each converge to zero independent of $\underset{\sim}{\psi}_{1n}$ as long as Conditions A.2.1 and A.3.1 are true.

For the first term write $\underset{\sim}{T}_1^{-1} = \underset{\sim}{\Sigma}_0^{-1} + \underset{\sim}{\Delta}$. Since $\operatorname{tr} \underset{\sim}{C} + \operatorname{tr} \underset{\sim}{D} = \operatorname{tr}(\underset{\sim}{C} + \underset{\sim}{D})$ for any matrices $\underset{\sim}{C}$ and $\underset{\sim}{D}$, write ϕ_0 as three terms and bound each separately. Each is of the form $\frac{1}{2n_i n_j} \left| \operatorname{tr} \underset{\sim}{E}_1 \underset{\sim}{G}_i \underset{\sim}{E}_2 \underset{\sim}{G}_j \right|$ and hence Propositions A.3.8 and A.3.9 apply to give the following bounds.

Table A.4.4.4

Bounds Used to Demonstrate Convergence to Zero of ϕ_0

Term	$\underset{\sim}{E}_1$	$\underset{\sim}{E}_2$	Bound for $\frac{1}{2n_in_j}\lvert \mathrm{tr}\ \underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j\rvert$
1	$\underset{\sim}{\Sigma}_0^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$	$\frac{\min(m_i,m_j)}{n_in_j}\ \frac{b}{\min\limits_{k=0,1,\ldots,p_1}(n_k\sigma_{0k})\sigma_{0i}\sigma_{0j}}$
2	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$	$\underset{\sim}{\Sigma}_0^{-1}$	$\frac{\min(m_i,m_j)}{n_in_j}\ \frac{b}{\min\limits_{k=0,1,\ldots,p_1}(n_k\sigma_{0k})\sigma_{0i}\sigma_{0j}}$
3	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$	$\frac{\min(m_i,m_j)}{n_in_j}\ \frac{b^2}{\min\limits_{k=0,1,\ldots,p_1}(n_k\sigma_{0k})^2\sigma_{0i}\sigma_{0j}}$

Again Proposition A.3.10 guarantees that all these bounds converge to zero independent of $\underset{\sim}{\psi}_{1n}$ and hence $\phi_0 \to 0$ independent of $\underset{\sim}{\psi}_{1n}$.

Handling ϕ_1 is fairly messy. Lemma B.16 applies giving four terms:

$$\phi_{11} = \frac{1}{2n_in_j}\left|(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'[\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_1^{-1} - \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j\underset{\sim}{\Sigma}_0^{-1}](\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)\right|,$$

$$\phi_{12} = \frac{1}{n_in_j}\left|\frac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)}{n_{p_1+1}}\underset{\sim}{X}'\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_1^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha})\right|,$$

$$\phi_{13} = \frac{1}{n_i n_j} \left| (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha}_0)' \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_i \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_j \underset{\sim}{T}_1^{-1} \underset{\sim}{X} \frac{(\underset{\sim}{\beta}_0 - \underset{\sim}{\beta}_1)}{n_{p_1+1}} \right|$$

$$\phi_{14} = \frac{1}{2n_i n_j} \left| \frac{(\underset{\sim}{\beta}_0 - \underset{\sim}{\beta}_1)}{n_{p_1+1}} \underset{\sim}{X}' \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_i \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_j \underset{\sim}{T}_1^{-1} \underset{\sim}{X} \frac{(\underset{\sim}{\beta}_0 - \underset{\sim}{\beta}_1)}{n_{p_1+1}} \right| .$$

ϕ_{14} is easy to dispose of.

$$\phi_{14}^2 \le \frac{1}{2n_i^2 n_j^2} \left[(\underset{\sim}{\beta}_0 - \underset{\sim}{\beta}_1)'(\underset{\sim}{\beta}_0 - \underset{\sim}{\beta}_1) \right] \left[\frac{1}{n_{p_1+1}^2} \lambda_{\max}(\underset{\sim}{X}' \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{X}) \right]^2$$

$$\cdot \lambda_{\max}(\underset{\sim}{A}' \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_j \underset{\sim}{T}_1^{-1} \underset{\sim}{A}\underset{\sim}{A}' \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_i \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_j \underset{\sim}{T}_1^{-1} \underset{\sim}{A})$$

by Proposition A.3.4,

$$\le \frac{1}{2n_i^2 n_j^2} p_0^2 b^4 B^2 2^6 \frac{1}{\sigma_{0i}^2 \sigma_{0j}^2}$$

by Propositions A.3.3, A.3.2, A.3.1, and A.3.10. This certainly converges to zero independent of $\underset{\sim}{\psi}_{1n}$.

Now use Proposition A.3.4 again with $\underset{\sim}{F}_1 = \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_i \underset{\sim}{T}_1^{-1}$ (which is symmetric) and $\underset{\sim}{F}_2 = \underset{\sim}{G}_j \underset{\sim}{T}_2^{-1}$ to obtain

$$\phi_{12}^2 \le \frac{1}{n_i^2 n_j^2} (\underset{\sim}{\beta}_0 - \underset{\sim}{\beta}_1)'(\underset{\sim}{\beta}_0 - \underset{\sim}{\beta}_1) \left[\frac{1}{n_{p_1+1}^2} \lambda_{\max}(\underset{\sim}{X}' \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{X}) \right] \lambda_{\max}(\underset{\sim}{A}' \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_i \underset{\sim}{T}_1^{-1} \underset{\sim}{A})^2$$

$$\cdot (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha}_0)' \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_j \underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1} \underset{\sim}{G}_j \underset{\sim}{T}_1^{-1} (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha}_0)$$

$$\leq \frac{1}{n_i^2} p_0 b^2 B \cdot \left(\frac{4}{\sigma_{0i}}\right)^2 (\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)' \left[\frac{\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_j}{n_j} \right] \underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1} \left[\frac{\underset{\sim}{G}_j\underset{\sim}{T}_1^{-1}}{n_j} \right] (\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0).$$

Since the first part clearly goes to zero it is sufficient to bound $(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'\underset{\sim}{F}_2'\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{F}_2(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$ for $\underset{\sim}{F}_2 = \frac{1}{n_j} \underset{\sim}{G}_j\underset{\sim}{T}_1^{-1}$. But since Conditions A.2.1 and A.3.1 are true, the remarks following Proposition A.3.7 show that all the necessary bounds were obtained while working on the terms for $\frac{\partial^2\lambda}{\partial\tau_j\partial\underset{\sim}{\beta}}$; these bounds are contained in Table A.4.4.3. This all guarantees that $\phi_{12} \to 0$ as $n \to \infty$ independent of $\underset{\sim}{\psi}_{1n}$ as long as Conditions A.2.1 and A.3.1 are true. ϕ_{13} obviously follows the same bounds as ϕ_{12} and hence also converges to zero.

To handle ϕ_{11} use the fact that $\underset{\sim}{T}_1^{-1} = \underset{\sim}{\Sigma}_0^{-1} + \underset{\sim}{\Delta}$ and hence $\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_1^{-1} - \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j\underset{\sim}{\Sigma}_0^{-1}$ breaks into seven terms of which two typical ones are $\underset{\sim}{\Delta}\underset{\sim}{G}_i\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{G}_j\underset{\sim}{\Sigma}_0^{-1}$ and $\underset{\sim}{\Delta}\underset{\sim}{G}_i\underset{\sim}{\Delta}\underset{\sim}{G}_j\underset{\sim}{\Sigma}_0^{-1}$. Each of the seven terms that ϕ_{11} breaks into are then of the form $\frac{1}{2}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'\underset{\sim}{F}_0\underset{\sim}{F}_1\underset{\sim}{F}_2(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$ for the following values of $\underset{\sim}{F}_0$, $\underset{\sim}{F}_1$, $\underset{\sim}{F}_2$ (again the n_i and n_j are inserted in the appropriate places).

Table A.4.4.5

Division of ϕ_{11} into Seven Terms with Appropriate Choices of $\underset{\sim}{F}_0, \underset{\sim}{F}_1$, and $\underset{\sim}{F}_2$

Term	$\underset{\sim}{F}_0$	$\underset{\sim}{F}_1$	$\underset{\sim}{F}_2$
1	$\frac{1}{n_i} \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$	$\frac{1}{n_j} \underset{\sim}{G}_j \underset{\sim}{\Sigma}_0^{-1}$
2	$\frac{1}{n_i} \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i$	$\frac{1}{n_j} \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j \underset{\sim}{T}_1^{-1}$	$(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$
3	$\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)$	$\frac{1}{n_i} \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1}$	$\frac{1}{n_j} \underset{\sim}{G}_j \underset{\sim}{\Sigma}_0^{-1}$
4	$\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)$	$\frac{1}{n_i n_j} \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_i \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j \underset{\sim}{T}_1^{-1}$	$(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$
5	$\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)$	$\frac{1}{n_i} \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_i \underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$	$\frac{1}{n_j} \underset{\sim}{G}_j \underset{\sim}{\Sigma}_0^{-1}$
6	$\frac{1}{n_i} \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_i$	$\frac{1}{n_j} \underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j \underset{\sim}{T}_1^{-1}$	$(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$
7	$\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)$	$\frac{1}{n_i n_j} \underset{\sim}{T}_1^{-1} \underset{\sim}{G}_i \underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{G}_j \underset{\sim}{T}_1^{-1}$	$(\underset{\sim}{\Sigma}_0 - \underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}$

By inspection of Table A.4.4.5, it is easily seen that all $\underset{\sim}{F}_0$ and $\underset{\sim}{F}_2$ are of such a form as to fit into Table A.4.4.3. Furthermore all $\underset{\sim}{F}_1$ are such that $\lambda_{max}(\underset{\sim}{A}'\underset{\sim}{F}_1'\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A}) \to 0$. (This is done by application of Lemma B.13 and Proposition A.3.2 and then using Table A.4.4.2.) These two facts together yield that each of the seven terms converges to zero and hence that $\phi_{11} \to 0$ as $n \to \infty$ independent of $\underset{\sim}{\psi}_{1n}$.

This now covers all possible cases of $\dfrac{\partial^2\lambda}{\partial\psi_i\partial\psi_j}$ and thus the lemma is proved. |||

A.4.5. Verification of Conditions 3.3.1.v and 3.3.1.i--Uniform Continuity of $\left.\dfrac{\partial^2\lambda}{\partial\underset{\sim}{\psi}\partial\underset{\sim}{\psi}'}\right|_{\underset{\sim}{\psi}_n}$ for $\underset{\sim}{\psi}_n \in S_b(\underset{\sim}{\psi}_{0n})$ in Probability

LEMMA A.4.5. For $\underset{\sim}{\psi}$ and $\lambda(\underset{\sim}{y},\underset{\sim}{\psi})$ as defined in Section 4.5 and for any $b > 0$, if Conditions A.2.1 and A.3.1 are true, then

$\dfrac{\partial^2\lambda(\underset{\sim}{y},\underset{\sim}{\psi})}{\partial\psi_i\partial\psi_j}$ is a uniformly continuous function of $\underset{\sim}{\psi}_n$ in $S_b(\underset{\sim}{\psi}_{0n})$,

$i,j=1,2,\ldots,p$.

PROOF.

Let $\eta > 0$ be given and let $\underset{\sim}{\psi}_{1n} \in S_b(\underset{\sim}{\psi}_{0n})$. It must be shown that there exists $\delta > 0$ (without loss of generality $\delta < \frac{b}{2}$) such that for

$\underset{\sim}{\psi}_{2n} \in S_\delta(\underset{\sim}{\psi}_{1n})$

$$\left| \frac{\partial^2 \lambda(\underset{\sim}{Y}, \underset{\sim}{\psi})}{\partial \psi_i \partial \psi_j} \Big|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{2n}} - \frac{\partial^2 \lambda(\underset{\sim}{Y}, \underset{\sim}{\psi})}{\partial \psi_i \partial \psi_j} \Big|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{1n}} \right| < \eta$$

for all i,j=1,2,...,p. Clearly it is sufficient to bound the various parts into which these expressions may break up by a constant times a function of δ which decreases as δ decreases. As in previous lemmae

$\underset{\sim}{\psi}_{an} = (\underset{\sim}{\beta}'_a, \tau_{a0}, \tau_{a1}, \ldots, \tau_{ap_1})'$, $\underset{\sim}{T}_a = \sum_{i=0}^{p_1} \frac{\tau_{ai}}{n_i} \underset{\sim}{G}_i$, for a = 1,2. $\underset{\sim}{T}_0 = \underset{\sim}{\Sigma}_0$;

$\frac{\underset{\sim}{\beta}_0}{n_{p_1+1}} = \underset{\sim}{\alpha}_0$.

The derivatives to be considered first are

$$\frac{\partial^2 \lambda}{\partial \underset{\sim}{\beta} \partial \underset{\sim}{\beta}'} \Big|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{an}} = \frac{1}{n^2_{p_1+1}} \underset{\sim}{X}' \underset{\sim}{T}_a^{-1} \underset{\sim}{X}, \quad a=0,1,2.$$

It is clearly sufficient for these derivatives to show that for any $\underset{\sim}{\xi}_1$, $\underset{\sim}{\xi}_2$, $p_0 \times 1$ such that $\underset{\sim}{\xi}'_1 \underset{\sim}{\xi}_1 = \underset{\sim}{\xi}'_2 \underset{\sim}{\xi}_2 = 1$, that $\phi = \frac{1}{n^2_{p_1+1}} |\underset{\sim}{\xi}'_1 \underset{\sim}{X}' (\underset{\sim}{T}_2^{-1} - \underset{\sim}{T}_1^{-1}) \underset{\sim}{X} \underset{\sim}{\xi}_2|$ can be bounded properly. But

$$\phi^2 \leq (\underset{\sim}{\xi}'_1 \underset{\sim}{\xi}_1)(\underset{\sim}{\xi}'_2 \underset{\sim}{\xi}_2) \left[\frac{1}{n^2_{p_1+1}} \lambda_{max}(\underset{\sim}{X}' \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{X})\right]^2 \lambda_{max} [\underset{\sim}{A}' (\underset{\sim}{T}_2^{-1} - \underset{\sim}{T}_1^{-1}) \underset{\sim}{A}]^2$$

by Proposition A.3.4 and the fact that $\underset{\sim}{T}_2^{-1} - \underset{\sim}{T}_1^{-1}$ is symmetric. But

$\underset{\sim}{T}_2^{-1} = \underset{\sim}{T}_1^{-1} + \underset{\sim}{T}_2^{-1} - \underset{\sim}{T}_1^{-1} = \underset{\sim}{T}_1^{-1} + \underset{\sim}{\Delta}_{21}$ where $\underset{\sim}{\Delta}_{21} = \underset{\sim}{T}_2^{-1}(\underset{\sim}{T}_1 - \underset{\sim}{T}_2)\underset{\sim}{T}_1^{-1}$, and $\underset{\sim}{\Delta}_{21}$ is symmetric; thus

$$\lambda_{\max}(\underset{\sim}{A}'(\underset{\sim}{T}_2^{-1} - \underset{\sim}{T}_1^{-1})\underset{\sim}{A})^2$$

$$= \lambda_{\max}(\underset{\sim}{A}'\underset{\sim}{\Delta}_{21}\underset{\sim}{A})^2$$

$$= \lambda_{\max}(\underset{\sim}{A}'\underset{\sim}{\Delta}_{21}\underset{\sim}{A}\underset{\sim}{A}'\Delta_{21}\underset{\sim}{A})$$

$$= \lambda_{\max}(\underset{\sim}{A}'\underset{\sim}{T}_1^{-1}\underset{\sim}{A}\underset{\sim}{A}^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{A}^{-t}\underset{\sim}{A}'\underset{\sim}{T}_2^{-1}\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{T}_2^{-1}\underset{\sim}{A}\underset{\sim}{A}^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{A}^{-t}\underset{\sim}{A}^{-1}\underset{\sim}{T}_1^{-1}\underset{\sim}{A})$$

$$\leq \lambda^2_{\max}(\underset{\sim}{T}_1^{-1}\underset{\sim}{\Sigma}_0)\lambda^2_{\max}(\underset{\sim}{T}_2^{-1}\underset{\sim}{\Sigma}_0)\max_{k=1,2,\ldots,n}|\lambda_k[\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)]|^2$$

by Lemmae B.8, B.11, B.8, B.14,

$$\leq 4\cdot 16 \cdot \frac{\delta^2}{\min\limits_{i=0,1,\ldots,p_1}(n_i\sigma_{0i})^2}$$

by Proposition A.3.2. Combining all the above, note that

$\left[\frac{1}{n^2_{p_1+1}}\lambda_{\max}(\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X})\right]^2 \leq B^2$ by Proposition A.3.10 and hence

$$\phi^2 \leq \frac{64B^2}{\min\limits_{i=0,1,\ldots,p_1}(n_i\sigma_{0i})^2}\;\delta^2 .$$

This is just of the form desired because there certainly exists a

constant such that $\dfrac{64B^2}{\min\limits_{i=0,1,\ldots,p_1} (n_i\sigma_{0i})^2}$ is less than that constant for all n.

For the next set of derivatives,

$$\frac{\partial^2\lambda}{\partial\tau_i\partial\underset{\sim}{\beta}}\Bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{an}} = \frac{1}{n_i n_{p_1+1}} \underset{\sim}{X}^1 \underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_a^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_a}{n_{p_1+1}}),\ i=0,1,\ldots,p_1,\ a=1,2.$$

Therefore it is sufficient to show for these derivatives that for all $\underset{\sim}{\xi}$ $p_0\times 1$ such that $\underset{\sim}{\xi}'\underset{\sim}{\xi} = 1$ that

$$\frac{1}{n_i n_{p_1+1}}\left|\underset{\sim}{\xi}'\underset{\sim}{X}'\left[\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\left(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_2}{n_{p_1+1}}\right)-\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\left(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_1}{n_{p_1+1}}\right)\right]\right|$$

is suitably bounded. Lemma B.15 applies here resulting in three terms to bound.

$$\phi_1 = \frac{1}{n_i n_{p_1+1}} |\underset{\sim}{\xi}'\underset{\sim}{X}'(\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1} - \underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1})(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)|,$$

$$\phi_2 = \frac{1}{n_i n_{p_1+1}} |\underset{\sim}{\xi}'\underset{\sim}{X}'(\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1} - \underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1})\underset{\sim}{X}\frac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)}{n_{p_1+1}}|,$$

$$\phi_3 = \frac{1}{n_i n_{p_1+1}} |\underset{\sim}{\xi}'\underset{\sim}{X}'\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{X}\frac{(\underset{\sim}{\beta}_1-\underset{\sim}{\beta}_2)}{n_{p_1+1}}| .$$

Consider ϕ_3 first.

$$\phi_3^2 \le \frac{1}{n_i^2} (\underset{\sim}{\xi}'\underset{\sim}{\xi})(\underset{\sim}{\beta}_1-\underset{\sim}{\beta}_2)'(\underset{\sim}{\beta}_1-\underset{\sim}{\beta}_2)\left[\frac{1}{n_{p_1+1}^2} \lambda_{max}(\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X})\right]^2 \lambda_{max}(\underset{\sim}{A}'\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{A})^2$$

by Proposition A.3.4,

$$\le \delta^2 \; \frac{1}{n_i^2} \cdot p_0 \cdot B^2 \cdot \frac{64}{\sigma_{0i}^2}$$

by Propositions A.3.1, A.3.2, A.3.10. This is again of the desired form.

For ϕ_2 as above $\underset{\sim}{T}_2^{-1} = \underset{\sim}{T}_1^{-1} + \underset{\sim}{\Delta}_{21}$ and thus there are three terms in the difference of the form $\underset{\sim}{\xi}'\underset{\sim}{X}'\underset{\sim}{F}_1\underset{\sim}{X} \dfrac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)}{n_{p_1+1}}$ which can be bounded by Propositions A.3.4, A.3.2, A.3.3, A.3.1 and A.3.10 as follows in Table A.4.5.1.

Table A.4.5.1

Division of ϕ_2 into Three Terms with Appropriate

$\underset{\sim}{F}_1$ and Bounds for Squares of Each Term

Term	$\underset{\sim}{F}_1$	Eventual Bound for its Square
1	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}$	$\dfrac{256p_0B^2}{n_i^2\sigma_{0i}^2 \min\limits_{j=0,1,\ldots,p_1} (n_j\sigma_{0j})^2} \cdot \delta$
2	$\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_1^{-1}$	$\dfrac{256p_0B^2}{n_i^2\sigma_{0i} \min\limits_{j=0,1,\ldots,p_1} (n_j\sigma_{0j})^2} \cdot \delta^2$
3	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_1^{-1}$	$\dfrac{4096p_0B^2}{n_i^2\sigma_{0i}^2 \min\limits_{j=0,1,\ldots,p_1} (n_j\sigma_{0j})^4} \cdot \delta^4$

All of these bounds have the desired form.

It remains to dispose of ϕ_1. Again there will be three terms of the form $\underset{\sim}{\xi}' \underset{\sim}{X}\, \underset{\sim}{F}_1\underset{\sim}{F}_2(\underset{\sim}{y}-\underset{\sim}{X}\,\underset{\sim}{\alpha}_0)$ which can be bounded by Proposition A.3.4.

Table A.4.5.2

Division of ϕ_1 into Three Terms with Appropriate

$\underset{\sim}{F}_1$ and $\underset{\sim}{F}_2$

Term	$\underset{\sim}{F}_1$	$\underset{\sim}{F}_2$
1	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\frac{1}{n_i}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}$
2	$\frac{1}{n_i}\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}$	$(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_1^{-1}$
3	$\frac{1}{n_i}\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}$	$(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_1^{-1}$

Again it is apparent that all these terms will yield proper bounds even after the further division into c^2 terms using the $\underset{\sim}{Q}_s$ matrices and $\underset{\sim}{w}_s$ vectors. For illustration consider Term 2 above.

$$\lambda_{max}(\underset{\sim}{A}'\underset{\sim}{F}_1'\underset{\sim}{A}\underset{\sim}{A}'\underset{\sim}{F}_1\underset{\sim}{A}) \le \frac{1}{n_i^2}\ \frac{64}{\sigma_{0i}^2}$$

by Propositions A.3.3 and A.3.2. Further, as in the comments following Proposition A.3.7, $F_2 = (T_1-T_2)T_1^{-1} = (\underset{\sim}{T}_1-\underset{\sim}{T}_2)(\underset{\sim}{\Sigma}_0^{-1}+\underset{\sim}{\Delta})$, $(\underset{\sim}{\Delta} = \underset{\sim}{T}_1^{-1}-\underset{\sim}{\Sigma}_0^{-1})$. Thus $\underset{\sim}{w}_s'\ \underset{\sim}{Q}_s'\ \underset{\sim}{A}'\ \underset{\sim}{F}_2'\ \underset{\sim}{A}^{-t}\ \underset{\sim}{A}^{-1}\ \underset{\sim}{F}_2\ \underset{\sim}{A}\ \underset{\sim}{Q}_s\ \underset{\sim}{w}_s$ breaks into four terms and only two need be bounded. The first term is

$\underset{\sim}{w}_s'\ \underset{\sim}{Q}_s'\ \underset{\sim}{A}'\ \underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{A}^{-t}\ \underset{\sim}{A}^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\ \underset{\sim}{\Sigma}_0^{-1}\ \underset{\sim}{A}\ \underset{\sim}{Q}\ \underset{\sim}{w}_s$; this is covered in

Proposition A.3.7 and has a bound involving δ^2 and the proper denominator. The second term is

$$\underset{\sim}{w}_s' \underset{\sim}{Q}_s' \underset{\sim}{A}' \underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_1^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{A}\, \underset{\sim}{Q}_s\, \underset{\sim}{w}_s$$

$$\le \underset{\sim}{w}_s' \underset{\sim}{Q}_s' \underset{\sim}{A}' \underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1}(\underset{\sim}{\Sigma}_0-\underset{\sim}{T}_1)\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{A}\, \underset{\sim}{Q}_s\, \underset{\sim}{w}_s$$

$$\cdot\ \lambda_{max}(\underset{\sim}{A}'\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{A}^{-t} \underset{\sim}{A}^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_1^{-1} \underset{\sim}{A})$$

by definition of characteristic root. But the first term is that of Proposition A.3.6 and is properly bounded and the second is less than or equal to $\dfrac{4}{\min\limits_{j=0,1,\ldots,p_1}(n_j\sigma_{0_j})} \cdot \delta^2$. All the other cases also give proper bounds. (All terms are bounded by a correct function of δ and a suitable constant.) This settles the case for ϕ_1 and hence for these derivatives of the form $\dfrac{\partial^2\lambda}{\partial\tau_i\partial\underset{\sim}{\beta}}$.

Now consider the derivatives of the form

$$\frac{\partial^2\lambda}{\partial\tau_i\partial\tau_j}\Bigg|_{\underset{\sim}{\psi}=\underset{\sim}{\psi}_{an}} = \frac{1}{2n_in_j}\Bigg[\Bigg(\mathrm{tr}\underset{\sim}{T}_a^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_a^{-1}\underset{\sim}{G}_j-2\Big(\underset{\sim}{y}-\underset{\sim}{X}\,\frac{\underset{\sim}{\beta}_a}{n_{p_1+1}}\Big)'\underset{\sim}{T}_a^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_a^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_a^{-1}\Big(\underset{\sim}{y}-\underset{\sim}{X}\,\frac{\underset{\sim}{\beta}_a}{n_{p_1+1}}\Big)\Bigg)\Bigg],$$

$$i,j=0,1,\ldots,p_1,\quad a=1,2.$$

It is sufficient for these derivatives to show that for all $i,j=0,1,\ldots,p_1$ that

$$\phi_0 = \frac{1}{2n_i n_j} \left|\mathrm{tr}\ \underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_j - \mathrm{tr}\ \underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_j\right|$$

and

$$\phi_1 = \frac{1}{n_i n_j} \left|\left(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_2}{n_{p_1+1}}\right)'\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_2^{-1}\left(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_2}{n_{p_1+1}}\right)\right.$$

$$\left. - \left(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_1}{n_{p_1+1}}\right)'\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_1^{-1}\left(\underset{\sim}{y}-\underset{\sim}{X}\frac{\underset{\sim}{\beta}_1}{n_{p_1+1}}\right)\right|$$

are both bounded by suitable functions of δ.

For ϕ_0 there are three terms of the form $\frac{1}{2n_i n_j}\left|\mathrm{tr}\ \underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j\right|$ and hence Propositions A.3.8, A.3.9 and A.3.10 apply to give the following bounds in Table A.4.5.3.

Table A.4.5.3

Division of ϕ_0 into Three Terms with Appropriate $\underset{\sim}{E}_1$ and $\underset{\sim}{E}_2$ and Bounds for Each Term

Term	$\underset{\sim}{E}_1$	$\underset{\sim}{E}_2$	Bound for $\frac{1}{2n_i n_j}\left\|\mathrm{tr}\ \underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j\right\|$
1	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}$	$\dfrac{8B}{\min\limits_{k=0,1,\ldots,p_1}(n_k\sigma_{0k})\sigma_{0i}\sigma_{0j}}\ \delta$
2	$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\dfrac{8B}{\min\limits_{k=0,1,\ldots,p_1}(n_k\sigma_{0k})\sigma_{0i}\sigma_{0j}}\ \delta$
3	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\dfrac{32B}{\min\limits_{k=0,1,\ldots,p_1}(n_k\sigma_{0k})^2\sigma_{0i}\sigma_{0j}}\ \delta^2$

All these bounds are of the proper form.

Now ϕ_1 breaks into nine parts by Lemma B.16 as follows:

$$\phi_{11} = \frac{1}{n_i n_j}\left|(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'(\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_2^{-1}-\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_1^{-1})(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)\right|$$

$$\phi_{12} = \frac{1}{n_i n_j}\left|\frac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)'}{n_{p_1+1}}\underset{\sim}{X}'(\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_2^{-1}-\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_1^{-1})(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)\right|$$

$$\phi_{13} = \frac{1}{n_i n_j}\left|(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'(\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_2^{-1}-\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_1^{-1})\underset{\sim}{X}\frac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)}{n_{p_1+1}}\right|$$

$$\phi_{14} = \frac{1}{n_i n_j}\left|\frac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)'}{n_{p_1+1}}\underset{\sim}{X}'(\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_2^{-1}-\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_1^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_1^{-1})\underset{\sim}{X}\frac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)}{n_{p_1+1}}\right|$$

$$\phi_{15} = \frac{1}{n_i n_j}\left|(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)'\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_2^{-1}X\frac{(\underset{\sim}{\beta}_1-\underset{\sim}{\beta}_2)}{n_{p_1+1}}\right|$$

$$\phi_{16} = \frac{1}{n_i n_j}\left|\frac{(\underset{\sim}{\beta}_1-\underset{\sim}{\beta}_2)'}{n_{p_1+1}}\underset{\sim}{X}'\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_2^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)\right|$$

$$\phi_{17} = \frac{1}{n_i n_j}\left|\frac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)'}{n_{p_1+1}}\underset{\sim}{X}'\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_2^{-1}\underset{\sim}{X}\frac{(\underset{\sim}{\beta}_1-\underset{\sim}{\beta}_2)}{n_{p_1+1}}\right|$$

$$\phi_{18} = \frac{1}{n_i n_j}\left|\frac{(\underset{\sim}{\beta}_1-\underset{\sim}{\beta}_2)'}{n_{p_1+1}}\underset{\sim}{X}'\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_2^{-1}\underset{\sim}{X}\frac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)}{n_{p_1+1}}\right|$$

$$\phi_{19} = \frac{1}{n_i n_j}\left|\frac{(\underset{\sim}{\beta}_1-\underset{\sim}{\beta}_2)'}{}\underset{\sim}{X}'\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_i\underset{\sim}{T}_2^{-1}\underset{\sim}{G}_j\underset{\sim}{T}_2^{-1}\underset{\sim}{X}\frac{(\underset{\sim}{\beta}_1-\underset{\sim}{\beta}_2)}{n_{p_1+1}}\right|$$

All of these terms can be properly bounded using Proposition A.3.4. The first four must be divided into seven terms in the usual way using $\underset{\sim}{T}_2^{-1} = \underset{\sim}{T}_1^{-1} + \underset{\sim}{\Delta}_{21}$. The last three are easily bounded as follows using Propositions A.3.4, A.3.2, and A.3.10.

$$\phi_{17}^2 \leq \frac{p_0^2 b^2}{n_i^2 n_j^2} \cdot B^2 \cdot \frac{4096}{\sigma_{0i}^2 \sigma_{0j}^2} \delta^2 ,$$

$$\phi_{18}^2 \leq \frac{p_0^2 b^2}{n_i^2 n_j^2} \quad B^2 \quad \frac{4096}{\sigma_{0i}^2 \sigma_{0j}^2} \delta^2 ,$$

$$\phi_{19}^2 \leq \frac{p_0^2}{n_i^2 n_j^2} \quad B^2 \quad \frac{4096}{\sigma_{0i}^2 \sigma_{0j}^2} \delta^4 .$$

All these bounds are of the proper form.

ϕ_{15} and ϕ_{16} are equal except for changing i and j so the same bound applies to each. Since a δ^2 term will come from the $(\underset{\sim}{\beta}_1 - \underset{\sim}{\beta}_2)$ part it must be verified only that boundedness for the remainder can be derived by Proposition A.3.4 with $\underset{\sim}{F}_1 = \frac{1}{n_i} \underset{\sim}{T}_2^{-1} \underset{\sim}{G}_i \underset{\sim}{T}_2^{-1}$ and $\underset{\sim}{F}_2 = \frac{1}{n_j} \underset{\sim}{G}_j \underset{\sim}{T}_2^{-1}$. But the same argument used above for $\underset{\sim}{F}_2 = \frac{1}{n_i} \underset{\sim}{G}_i \underset{\sim}{T}_1^{-1}$ can also be applied to this $\underset{\sim}{F}_2$ yielding a proper bound. Thus ϕ_{15} and ϕ_{16} can be properly bounded.

Now the decomposition of the other terms is summarized; the source of the all important δ terms is noted. It can be verified by inspection

that all bounds will be of proper form. ϕ_{14} breaks into seven terms of the form $\frac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)'}{n_{p_1+1}} \underset{\sim}{X}'\underset{\sim}{F}_1\underset{\sim}{X} \frac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)}{n_{p_1+1}}$ with $\underset{\sim}{F}_1 = \underset{\sim}{E}_1\underset{\sim}{G}_i\underset{\sim}{E}_2\underset{\sim}{G}_j\underset{\sim}{E}_3$. Such terms are bounded by Propositions A.3.4, A.3.3, A.3.2, and A.3.10.

Table A.4.5.4

Division of ϕ_{14} into Seven Terms with Appropriate $\underset{\sim}{E}_1$, $\underset{\sim}{E}_2$ and $\underset{\sim}{E}_3$ and Source of δ Term

Term	$\underset{\sim}{E}_1$	$\underset{\sim}{E}_2$	$\underset{\sim}{E}_3$	Source of δ Term
1	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{E}_1$
2	$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{E}_2$
3	$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{E}_3$
4	$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{E}_2,\underset{\sim}{E}_3$
5	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{E}_1,\underset{\sim}{E}_3$
6	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}$	$\underset{\sim}{E}_1,\underset{\sim}{E}_2$
7	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{T}_1^{-1}(\underset{\sim}{T}_1-\underset{\sim}{T}_2)\underset{\sim}{T}_2^{-1}$	$\underset{\sim}{E}_1,\underset{\sim}{E}_2,\underset{\sim}{E}_3$

ϕ_{12} and ϕ_{13} are the transpose of each other upon interchange of i and j. ϕ_{12} breaks into seven terms of the form $\frac{(\underset{\sim}{\beta}_0-\underset{\sim}{\beta}_1)'}{n_{p_1+1}} \underset{\sim}{X}'\underset{\sim}{F}_1\underset{\sim}{F}_2(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha}_0)$ which can be bounded by Propositions A.3.1-A.3.7 and A.3.10.

Table A.4.5.5

Division of ϕ_{12} into Seven Terms with Appropriate

$\mathbf{F}_1$ and $\mathbf{F}_2$ and Source of δ Term

Term	$\mathbf{F}_1$	$\mathbf{F}_2$	Source of δ Term
1	$\frac{1}{n_i}\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}\mathbf{G}_i\mathbf{T}_1^{-1}$	$\frac{1}{n_j}\mathbf{G}_j\mathbf{T}_1^{-1}$	$\mathbf{F}_1$
2	$\frac{1}{n_i}\mathbf{T}_1^{-1}\mathbf{G}_i\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}$	$\frac{1}{n_j}\mathbf{G}_j\mathbf{T}_1^{-1}$	$\mathbf{F}_1$
3	$\frac{1}{n_i n_j}\mathbf{T}_1^{-1}\mathbf{G}_i\mathbf{T}_1^{-1}\mathbf{G}_j\mathbf{T}_2^{-1}$	$(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_1^{-1}$	$\mathbf{F}_2$
4	$\frac{1}{n_i n_j}\mathbf{T}_1^{-1}\mathbf{G}_i\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}\mathbf{G}_j\mathbf{T}_2^{-1}$	$(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_1^{-1}$	$\mathbf{F}_1,\mathbf{F}_2$
5	$\frac{1}{n_i n_j}\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}\mathbf{G}_i\mathbf{T}_1^{-1}\mathbf{G}_j\mathbf{T}_2^{-1}$	$(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_1^{-1}$	$\mathbf{F}_1,\mathbf{F}_2$
6	$\frac{1}{n_i}\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}\mathbf{G}_i\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}$	$\frac{1}{n_j}\mathbf{G}_j\mathbf{T}_1^{-1}$	$\mathbf{F}_1$
7	$\frac{1}{n_i n_j}\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}\mathbf{G}_i\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}\mathbf{G}_j\mathbf{T}_2^{-1}$	$(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_1^{-1}$	$\mathbf{F}_1,\mathbf{F}_2$

ϕ_{11} breaks into seven terms of the form $(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha}_0)'\mathbf{F}_0\mathbf{F}_1\mathbf{F}_2(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha}_0)$ which can be bounded by Propositions A.3.1-A.3.7 and A.3.10.

Table A.4.5.6

Division of ϕ_{11} into Seven Terms with Appropriate

$\mathbf{F}_0$, $\mathbf{F}_1$ and $\mathbf{F}_2$ and Source of δ Term

Term	$\mathbf{F}_0$	$\mathbf{F}_1$	$\mathbf{F}_2$	Source of δ Term
1	$\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)$	$\frac{1}{n_i}\mathbf{T}_2^{-1}\mathbf{G}_i\mathbf{T}_1^{-1}$	$\frac{1}{n_j}\mathbf{G}_j\mathbf{T}_1^{-1}$	$\mathbf{F}_0$
2	$\frac{1}{n_i}\mathbf{T}_1^{-1}\mathbf{G}_i$	$\mathbf{T}_2^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}$	$\frac{1}{n_j}\mathbf{G}_j\mathbf{T}_1^{-1}$	$\mathbf{F}_1$
3	$\frac{1}{n_i}\mathbf{T}_1^{-1}\mathbf{G}_i$	$\frac{1}{n_j}\mathbf{T}_1^{-1}\mathbf{G}_j\mathbf{T}_2^{-1}$	$(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_1^{-1}$	$\mathbf{F}_2$
4	$\frac{1}{n_i}\mathbf{T}_1^{-1}\mathbf{G}_i$	$\frac{1}{n_j}\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}\mathbf{G}_j\mathbf{T}_2^{-1}$	$(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_1^{-1}$	$\mathbf{F}_1,\mathbf{F}_2$
5	$\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)$	$\frac{1}{n_i n_j}\mathbf{T}_2^{-1}\mathbf{G}_i\mathbf{T}_1^{-1}\mathbf{G}_j\mathbf{T}_2^{-1}$	$(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_1^{-1}$	$\mathbf{F}_0,\mathbf{F}_2$
6	$\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)$	$\frac{1}{n_i}\mathbf{T}_2^{-1}\mathbf{G}_i\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}$	$\frac{1}{n_j}\mathbf{G}_j\mathbf{T}_1^{-1}$	$\mathbf{F}_0,\mathbf{F}_1$
7	$\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)$	$\frac{1}{n_i n_j}\mathbf{T}_2^{-1}\mathbf{G}_i\mathbf{T}_1^{-1}(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_2^{-1}\mathbf{G}_j\mathbf{T}_2^{-1}$	$(\mathbf{T}_1-\mathbf{T}_2)\mathbf{T}_1^{-1}$	$\mathbf{F}_0,\mathbf{F}_1,\mathbf{F}_2$

As noted before every term above yields a bound of the proper form after all decompositions (including the $\mathbf{Q}_s\mathbf{w}_s$ decomposition) have been made. Each term also includes a δ term. Thus all terms have proper bounds and the lemma is true. |||

APPENDIX B

ALGEBRAIC LEMMAE USED IN PREVIOUS CHAPTERS

The following lemmae are used in previous chapters. No claim of originality is made for any of them. They are collected here so that the reader may easily refer to them. Only a few proofs are given. The other lemmae can be proved by algebraic manipulation or application of simple well known results. References are given where appropriate.

LEMMA B.1. *If $\underset{\sim}{y} \sim \mathcal{N}_n(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$ with $\underset{\sim}{\Sigma}$ positive definite and $\underset{\sim}{B}$ is an nxn symmetric constant matrix and $\underset{\sim}{b}$ an nx1 constant vector then*

$$\mathcal{E}(\underset{\sim}{y}-\underset{\sim}{b})'\underset{\sim}{B}(\underset{\sim}{y}-\underset{\sim}{b}) = \operatorname{tr} \underset{\sim}{B}\underset{\sim}{\Sigma} + (\underset{\sim}{\mu}-\underset{\sim}{b})'\underset{\sim}{B}(\underset{\sim}{\mu}-\underset{\sim}{b})$$

and

$$\operatorname{Var}(\underset{\sim}{y}-\underset{\sim}{b})'\underset{\sim}{B}(\underset{\sim}{y}-\underset{\sim}{b}) = 2\operatorname{tr}(\underset{\sim}{B}\underset{\sim}{\Sigma})^2 + 4(\underset{\sim}{\mu}-\underset{\sim}{b})'\underset{\sim}{B}\underset{\sim}{\Sigma}\underset{\sim}{B}(\underset{\sim}{\mu}-\underset{\sim}{b})$$

PROOF.

This lemma is easily proved by algebraic manipulation of the moments up to fourth order of the multivariate normal distribution. These moments can be found in Anderson (1958:39). |||

LEMMA B.2. *If $\underset{\sim}{z} \sim \mathcal{N}_n(\underset{\sim}{0}, \underset{\sim}{I})$ and $\underset{\sim}{b}$ is an nx1 constant vector and $\underset{\sim}{\Lambda}$ is an nxn diagonal constant matrix, then*

$$\mathcal{E}\left\{e^{i(\underset{\sim}{b}'\underset{\sim}{z}+\underset{\sim}{z}'\underset{\sim}{\Lambda}\underset{\sim}{z})}\right\} = |\underset{\sim}{I}-2i\underset{\sim}{\Lambda}|^{-\frac{1}{2}}\, e^{-\frac{1}{2}\underset{\sim}{b}'(\underset{\sim}{I}-2i\underset{\sim}{\Lambda})^{-1}\underset{\sim}{b}}.$$

PROOF.

The proof of this lemma is merely an extension of proofs of the characteristic function of a multivariate normal random variable. It is done by completing the square. A reference is Plackett (1960:16-17).

LEMMA B.3. <u>If</u> $x_1, x_2, \ldots, x_p$ <u>are all nonnegative and at least one</u> x_i <u>is positive and</u> $b_1, b_2, \ldots, b_p$ <u>are all positive and</u> $a_1, a_2, \ldots, a_p$ <u>have any sign,</u>

$$\min_{i=1,2,\ldots,p} \frac{a_i}{b_i} \le \frac{\sum_{i=1}^{p} x_i a_i}{\sum_{i=1}^{p} x_i b_i} \le \max_{i=1,2,\ldots,p} \frac{a_i}{b_i} .$$

PROOF.

Let $\rho_i \equiv \dfrac{x_i b_i}{\sum_{j=1}^{p} x_j b_j}$. ($\sum_{j=1}^{p} x_j b_j$ is positive because at least one x_j is positive and all the b_j are.) Then $0 \le \rho_i \le 1$ and $\sum_{i=1}^{p} \rho_i = 1.$

Let $c_i = \dfrac{a_i}{b_i}$. Then $\sum_{i=1}^{p} \rho_i c_i$ is a weighted arithmetic mean of the c_i and is thus between $\min c_i = \min \dfrac{a_i}{b_i}$ and $\max c_i = \max \dfrac{a_i}{b_i}$. But

$$\sum_{i=1}^{p} \rho_i c_i = \sum_{i=1}^{p} \frac{x_i b_i}{\sum_{j=1}^{p} x_j b_j} \cdot \frac{a_i}{b_i} = \frac{\sum_{i=1}^{p} x_i a_i}{\sum_{i=1}^{p} x_i b_i}$$

and the lemma is proved. |||

LEMMA B.4. If $\underset{\sim}{A}$ is n×n positive definite and $\underset{\sim}{B}$ is n×n symmetric and if $\lambda_1(\underset{\sim}{C}) \geq \lambda_2(\underset{\sim}{C}) \geq \ldots \geq \lambda_n(\underset{\sim}{C})$ are the characteristic roots of any n×n matrix $\underset{\sim}{C}$, and if $\underset{\sim}{\alpha}_1, \ldots, \underset{\sim}{\alpha}_n$ are any n×1 vectors, then

$$\lambda_i(\underset{\sim}{A}^{-1}\underset{\sim}{B}) \leq \sup_{\substack{\underset{\sim}{x} \neq \underset{\sim}{0} \\ \underset{\sim}{x}'\underset{\sim}{\alpha}_j = 0 \\ j=1,2,\ldots,i-1}} \frac{\underset{\sim}{x}'\underset{\sim}{B}\underset{\sim}{x}}{\underset{\sim}{x}'\underset{\sim}{A}\underset{\sim}{x}}$$

and

$$\lambda_i(\underset{\sim}{A}^{-1}\underset{\sim}{B}) \geq \inf_{\substack{\underset{\sim}{x} \neq \underset{\sim}{0} \\ \underset{\sim}{x}'\underset{\sim}{\alpha}_j = 0 \\ j=1,2,\ldots,n-i}} \frac{\underset{\sim}{x}'\underset{\sim}{B}\underset{\sim}{x}}{\underset{\sim}{x}'\underset{\sim}{A}\underset{\sim}{x}} .$$

PROOF.

The proof of the first statement is given in Anderson and Das Gupta (1963). The second statement is proved analogously. |||

LEMMA B.5. If $\underset{\sim}{A}$ is n×n positive definite and $\underset{\sim}{B}$ is n×n symmetric then

$$\lambda_i(\underset{\sim}{A}^{-1}\underset{\sim}{B}) \geq \inf_{\underset{\sim}{x} \in \mathcal{L}} \frac{\underset{\sim}{x}'\underset{\sim}{B}\underset{\sim}{x}}{\underset{\sim}{x}'\underset{\sim}{A}\underset{\sim}{x}}$$

where $\mathcal{L}$ is a linear space of dimension i.

PROOF.

$\mathcal{L}$ has a basis of i vectors. Let them be $\underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2, \ldots, \underset{\sim}{\beta}_i$. This basis can be extended to a basis for R_n by adding n-i more vectors orthogonal

to $\underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2, \ldots, \underset{\sim}{\beta}_i$. Let these vectors be $\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, \ldots, \underset{\sim}{\alpha}_{n-i}$. Then $\underset{\sim}{x} \in \mathcal{L}$ if and only if $\underset{\sim}{x}'\underset{\sim}{\alpha}_j = 0$ for $j=1,2,\ldots,n-i$. Lemma B.4 then applies, proving this lemma. |||

LEMMA B.6. <u>For $\underset{\sim}{A}$ and $\underset{\sim}{B}$ as in Lemma B.4, $\lambda_i(\underset{\sim}{A}^{-1}\underset{\sim}{B}) = \dfrac{\underset{\sim}{x}_i'\underset{\sim}{B}\underset{\sim}{x}_i}{\underset{\sim}{x}_i'\underset{\sim}{A}\underset{\sim}{x}_i}$ where $\underset{\sim}{x}_i$ is the characteristic vector associated with λ_i.</u>

LEAMMA B.7. <u>Suppose $\underset{\sim}{B}_1 = \sum_{i=0}^{p} b_{1i}\underset{\sim}{G}_i$, $\underset{\sim}{B}_0 = \sum_{i=0}^{p} b_{0i}\underset{\sim}{G}_i$, all G_i are positive semidefinite and at least one is positive definite and</u> $b_{0i} > 0$, $i=0,1,\ldots,p$. <u>Then</u>

$$\text{i)}\quad \lambda_j(\underset{\sim}{B}_0^{-1}\underset{\sim}{B}_1) = \frac{\sum_{i=0}^{p} b_{1i}\underset{\sim}{x}_j'\underset{\sim}{G}_i\underset{\sim}{x}_j}{\sum_{i=0}^{p} b_{0i}\underset{\sim}{x}_j'\underset{\sim}{G}_i\underset{\sim}{x}_j}$$

<u>where $\underset{\sim}{x}_j$ is the characteristic vector associated with</u> λ_j,

$$\text{ii)}\quad \lambda_1(\underset{\sim}{B}_0^{-1}\underset{\sim}{B}_1) \le \max_{i=0,1,\ldots,p} \frac{b_{1i}}{b_{0i}} .$$

iii) <u>If $\underset{\sim}{H}$ is an</u> $n\times m$ <u>matrix such that $\underset{\sim}{x}'\underset{\sim}{H} = \underset{\sim}{0}$ implies</u> $\underset{\sim}{x}'\underset{\sim}{G}_i\underset{\sim}{x} = 0$, $i=j+1,\ldots,p$, <u>for some</u> $j=0,1,\ldots,p$ <u>then</u>

$$\lambda_{m+1}(\underset{\sim}{B}_0^{-1}\underset{\sim}{B}_1) \le \max_{i=0,1,\ldots,j} \frac{b_{1i}}{b_{0i}} .$$

PROOF.

Statement i) follows from Lemma B.6 and the definitions of $\underset{\sim}{B}_0$ and $\underset{\sim}{B}_1$. Statement ii) is a simple application of Lemma B.3 to part i). To prove Statement iii), observe that

$$\lambda_{m+1}(\underset{\sim}{B}_0^{-1}\underset{\sim}{B}_1) \leq \sup_{\substack{\underset{\sim}{x} \neq \underset{\sim}{0} \\ \underset{\sim}{x}'\underset{\sim}{H}=\underset{\sim}{0}}} \frac{\underset{\sim}{x}'\underset{\sim}{B}_1\underset{\sim}{x}}{\underset{\sim}{x}'\underset{\sim}{B}_0\underset{\sim}{x}}$$

by Lemma B.4,

$$\leq \sup_{\underset{\sim}{x} \neq \underset{\sim}{0}} \frac{\sum_{i=0}^{j} b_{1i}\underset{\sim}{x}'\underset{\sim}{G}_i\underset{\sim}{x}}{\sum_{i=0}^{j} b_{0i}\underset{\sim}{x}'\underset{\sim}{G}_i\underset{\sim}{x}}$$

$$\leq \max_{i=0,1,\ldots,j} \frac{b_{1i}}{b_{0i}}$$

by Lemma B.3. Note that in both parts ii) and iii) the correspondences between Lemma B.3 and Lemma B.7 are as follows:

Lemma B.3	Lemma B.7
x_i	$\underset{\sim}{x}'\underset{\sim}{G}_i\underset{\sim}{x}$
a_i	b_{1i}
b_i	b_{0i} . |||

LEMMA B.8. Let $\underset{\sim}{B}$ be an $n \times n$ symmetric matrix and $\underset{\sim}{C}$ an $n \times m$ matrix, then

$$\max_{j=1,2,\ldots,m} |\lambda_j(\underset{\sim}{C}'\underset{\sim}{B}\underset{\sim}{C})| \le \lambda_{\max}(\underset{\sim}{C}'\underset{\sim}{C}) \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{B})|.$$

PROOF.

$$\lambda_j(\underset{\sim}{C}'\underset{\sim}{B}\underset{\sim}{C}) = \frac{\underset{\sim}{x}_j'\underset{\sim}{C}'\underset{\sim}{B}\underset{\sim}{C}\underset{\sim}{x}_j}{\underset{\sim}{x}_j'\underset{\sim}{x}_j}$$

$$= \frac{\underset{\sim}{x}_j'\underset{\sim}{C}'\underset{\sim}{B}\underset{\sim}{C}\underset{\sim}{x}_j}{\underset{\sim}{x}_j'\underset{\sim}{C}'\underset{\sim}{C}\underset{\sim}{x}_j} \cdot \frac{\underset{\sim}{x}_j'\underset{\sim}{C}'\underset{\sim}{C}\underset{\sim}{x}_j}{\underset{\sim}{x}_j'\underset{\sim}{x}_j},$$

where $\underset{\sim}{x}_j$ is the characteristic vector associated with λ_j. (If $\underset{\sim}{C}\underset{\sim}{x}_j = 0$ that root is not of interest in any case since only the nonzero characteristic roots of $\underset{\sim}{C}'\underset{\sim}{B}\underset{\sim}{C}$ are of concern. Therefore, since $\underset{\sim}{x}_j'\underset{\sim}{C}'\underset{\sim}{C}\underset{\sim}{x}_j \ge 0$

$$|\lambda_j(\underset{\sim}{C}'\underset{\sim}{B}\underset{\sim}{C})| = \left|\frac{\underset{\sim}{x}_j'\underset{\sim}{C}'\underset{\sim}{B}\underset{\sim}{C}\underset{\sim}{x}_j}{\underset{\sim}{x}_j'\underset{\sim}{C}'\underset{\sim}{C}\underset{\sim}{x}_j}\right| \frac{\underset{\sim}{x}_j'\underset{\sim}{C}'\underset{\sim}{C}\underset{\sim}{x}_j}{\underset{\sim}{x}_j'\underset{\sim}{x}_j}$$

$$\le \sup_{\underset{\sim}{y} \ne \underset{\sim}{0}} \left|\frac{\underset{\sim}{y}'\underset{\sim}{B}\underset{\sim}{y}}{\underset{\sim}{y}'\underset{\sim}{y}}\right| \lambda_{\max}(\underset{\sim}{C}'\underset{\sim}{C})$$

$$= \lambda_{\max}(\underset{\sim}{C}'\underset{\sim}{C}) \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{B})|. \quad |||$$

LEMMA B.9. Let $\underset{\sim}{A}$ and $\underset{\sim}{B}$ be $n \times n$ matrices with $\underset{\sim}{A}$ nonsingular. Then $\lambda_i(\underset{\sim}{A}\underset{\sim}{B}) = \lambda_i(\underset{\sim}{B}\underset{\sim}{A})$.

LEMMA B.10. Let $\underset{\sim}{B}$ be an n×n matrix. Then $\lambda_i(\underset{\sim}{B}) = \lambda_i(\underset{\sim}{B}')$.

LEMMA B.11. Let $\underset{\sim}{A}$ be an m×n matrix and $\underset{\sim}{B}$ be an n×m matrix. Then the nonzero characteristic roots of $\underset{\sim}{AB}$ are equal to the nonzero characteristic roots of $\underset{\sim}{BA}$.

LEMMA B.12. Let $\underset{\sim}{x}$ and $\underset{\sim}{y}$ be n×1 vectors and $\underset{\sim}{A}$ an n×n matrix. Then $(\underset{\sim}{x}'\underset{\sim}{A}\underset{\sim}{y})^2 \le (\underset{\sim}{x}'\underset{\sim}{x})(\underset{\sim}{y}'\underset{\sim}{y})\ \lambda_{max}(\underset{\sim}{A}'\underset{\sim}{A})$.

LEMMA B.13. Let $\underset{\sim}{A}$ be n×n positive semidefinite and let $\underset{\sim}{B}$ be n×n. Then

$$\max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{AB})| \le \lambda_{max}(\underset{\sim}{A}) \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{B})|.$$

PROOF.

This lemma follows immediately from Lemmae B.8 and B.9 and the fact that there exists $\underset{\sim}{C}$ such that $\underset{\sim}{A}=\underset{\sim}{C}'\underset{\sim}{C}$. |||

LEMMA B.14. For $\underset{\sim}{A}$ n×n, $\max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{A}^j)| = \max_{k=1,2,\ldots,n} |\lambda_k(\underset{\sim}{A})|^j$, $j=1,2,3,\ldots$.

LEMMA B.15. If $\underset{\sim}{S}_1$ and $\underset{\sim}{S}_2$ are n×n matrices and $\underset{\sim}{t}_0$, $\underset{\sim}{t}_1$, $\underset{\sim}{t}_2$ and $\underset{\sim}{y}$ are n×1 vectors, then the following statements are true.

$$\text{i)}\quad \underset{\sim}{S}_2(\underset{\sim}{y}-\underset{\sim}{t}_2) - \underset{\sim}{S}_1(\underset{\sim}{y}-\underset{\sim}{t}_1) = (\underset{\sim}{S}_2-\underset{\sim}{S}_1)(\underset{\sim}{y}-\underset{\sim}{t}_0) + (\underset{\sim}{S}_2-\underset{\sim}{S}_1)(\underset{\sim}{t}_0-\underset{\sim}{t}_1) + \underset{\sim}{S}_2(\underset{\sim}{t}_1-\underset{\sim}{t}_2) .$$

ii) $(\mathbf{y}-\mathbf{t}_2)'\mathbf{S}_2(\mathbf{y}-\mathbf{t}_2) - (\mathbf{y}-\mathbf{t}_1)'\mathbf{S}_1(\mathbf{y}-\mathbf{t}_1)$

$$= (\mathbf{y}-\mathbf{t}_0)'(\mathbf{S}_2-\mathbf{S}_1)(\mathbf{y}-\mathbf{t}_0)+(\mathbf{t}_0-\mathbf{t}_1)'(\mathbf{S}_2-\mathbf{S}_1)(\mathbf{y}-\mathbf{t}_0)$$

$$+ (\mathbf{y}-\mathbf{t}_0)'(\mathbf{S}_2-\mathbf{S}_1)(\mathbf{t}_0-\mathbf{t}_1)+(\mathbf{t}_0-\mathbf{t}_1)'(\mathbf{S}_2-\mathbf{S}_1)(\mathbf{t}_0-\mathbf{t}_1)$$

$$+ (\mathbf{y}-\mathbf{t}_0)'\mathbf{S}_2(\mathbf{t}_1-\mathbf{t}_2)+(\mathbf{t}_1-\mathbf{t}_2)'\mathbf{S}_2(\mathbf{y}-\mathbf{t}_0)$$

$$+ (\mathbf{t}_0-\mathbf{t}_1)'\mathbf{S}_2(\mathbf{t}_1-\mathbf{t}_2)+(\mathbf{t}_1-\mathbf{t}_2)'\mathbf{S}_2(\mathbf{t}_0-\mathbf{t}_1)$$

$$+ (\mathbf{t}_1-\mathbf{t}_2)'\mathbf{S}_2(\mathbf{t}_1-\mathbf{t}_2) \ .$$

LEMMA B.16. _If $\mathbf{S}_1$ and $\mathbf{S}_0$ are nxn matrices and $\mathbf{t}_0$, $\mathbf{t}_1$ and $\mathbf{y}$ are nx1 vectors, then the following statements are true._

i) $\mathbf{S}_1(\mathbf{y}-\mathbf{t}_1) - \mathbf{S}_0(\mathbf{y}-\mathbf{t}_0) = (\mathbf{S}_1-\mathbf{S}_0)(\mathbf{y}-\mathbf{t}_0) + \mathbf{S}_1(\mathbf{t}_0-\mathbf{t}_1)$.

ii) $(\mathbf{y}-\mathbf{t}_1)\mathbf{S}_1(\mathbf{y}-\mathbf{t}_1) - (\mathbf{y}-\mathbf{t}_0)\mathbf{S}_0(\mathbf{y}-\mathbf{t}_0)$

$$= (\mathbf{y}-\mathbf{t}_0)'(\mathbf{S}_1-\mathbf{S}_0)(\mathbf{y}-\mathbf{t}_0)$$

$$+ (\mathbf{t}_0-\mathbf{t}_1)'\mathbf{S}_1(\mathbf{y}-\mathbf{t}_0)$$

$$+ (\mathbf{y}-\mathbf{t}_0)'\mathbf{S}_1(\mathbf{t}_0-\mathbf{t}_1)$$

$$+ (\mathbf{t}_0-\mathbf{t}_1)'\mathbf{S}_1(\mathbf{t}_0-\mathbf{t}_1) \ .$$

APPENDIX C

A COMPUTER PROGRAM TO IMPLEMENT THE ITERATIVE PROCEDURE

C.1. Algebra Used in the Computer Program

A computer program has been written to implement The Iterative Procedure, which was introduced in Chapter 5. This section contains a brief description of the algebra used to write the program. This algebraic manipulation was used instead of more straightforward methods of calculation in order to reduce core storage requirements in the computer. The straightforward methods require the storage of several n×n matrices in order to compute the quantities required for the solution of the iterative equations ($\underset{\sim}{\Sigma}$ and perhaps each of the $\underset{\sim}{G}_i$). Since this requires a great deal of core for even moderate n, some improvement is needed. The algebraic manipulations given here reduce the maximum dimension of the matrices which must be stored to $m \equiv \Sigma_{i=1}^{p_1} m_i$. This is usually much smaller than n. For example, in the two-way balanced layout described in Section 6.1, n=IJK and m=IJ+I+J; even for small I,J,K appreciable savings can result. For instance, if I,J,K is 2,3,3 n=18 but m=11 and if I,J,K is 3,6,4 n=72 but m=27. The basic result used in these manipulations is due to Woodbury (1950) and is stated as Proposition C.1.1.

PROPOSITION C.1.1. <u>Let</u> $\underset{\sim}{\Sigma} = \sigma_0 \underset{\sim}{I} + \underset{\sim}{U}\underset{\sim}{D}\underset{\sim}{U}'$, <u>where</u> $\underset{\sim}{\Sigma}$ <u>and</u> $\underset{\sim}{I}$ <u>are</u> n×n <u>and</u> $\underset{\sim}{U}$ <u>is</u>

n×m and $\underset{\sim}{D}$ is m×m diagonal and nonsingular. Then

$$\underset{\sim}{\Sigma}^{-1} = \frac{1}{\sigma_0}\left(\underset{\sim}{I}-\underset{\sim}{U}(\sigma_0\underset{\sim}{D}^{-1}+\underset{\sim}{U}'\underset{\sim}{U})^{-1}\underset{\sim}{U}'\right).$$

PROOF: This proposition is a special case of a result of Woodbury (1950). |||

The matrix which must be inverted for this computer program may be written in the above form, where $\underset{\sim}{U} = [\underset{\sim}{U}_1:\underset{\sim}{U}_2:\ldots:\underset{\sim}{U}_{p_1}]$ and

$$\underset{\sim}{D} = \begin{bmatrix} \underset{\sim}{D}_1 & \underset{\sim}{0} & \cdots & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{D}_2 & & \underset{\sim}{0} \\ \vdots & \vdots & & \vdots \\ \underset{\sim}{0} & \underset{\sim}{0} & & \underset{\sim}{D}_{p_1} \end{bmatrix},$$

where $\underset{\sim}{D}_i = \sigma_i \underset{\sim}{I}_{m_i}$. Thus $m = \Sigma_{i=1}^{p_1} m_i$. Let $\underset{\sim}{F} \equiv \underset{\sim}{U}'\underset{\sim}{U}$ be partitioned as

$$\underset{\sim}{F} = \begin{bmatrix} \underset{\sim}{F}_{11} & \cdots & \underset{\sim}{F}_{1p_1} \\ \vdots & & \vdots \\ \underset{\sim}{F}_{p_1 1} & \cdots & \underset{\sim}{F}_{p_1 p_1} \end{bmatrix},$$

where $\underset{\sim}{F}_{ij} = \underset{\sim}{U}_i'\underset{\sim}{U}_j$. Let $\underset{\sim}{E} \equiv (\sigma_0\underset{\sim}{D}^{-1} + \underset{\sim}{F})$ be partitioned similarly. Let

$$\underset{\sim}{w}_0 = \underset{\sim}{U}'\underset{\sim}{y} = \begin{bmatrix} \underset{\sim}{w}_1^{(0)} \\ \underset{\sim}{w}_2^{(0)} \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{w}_{p_1}^{(0)} \end{bmatrix},$$

where $\underset{\sim}{w}_i^{(0)} = \underset{\sim}{U}_i'\underset{\sim}{y}$. Let

$$\underset{\sim}{H} = \underset{\sim}{U}'\underset{\sim}{X} = \begin{bmatrix} \underset{\sim}{H}_1 \\ \underset{\sim}{H}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{H}_{p_1} \end{bmatrix}$$

where $\underset{\sim}{H}_i = \underset{\sim}{U}_i'\,\underset{\sim}{X}$. Let

$$\underset{\sim}{A}_0 = \underset{\sim}{X}'\underset{\sim}{X}$$

and

$$\underset{\sim}{b}_0 = \underset{\sim}{X}'\underset{\sim}{y}\ .$$

Then $\underset{\sim}{F}$ and $\underset{\sim}{E}$ are m×m, $\underset{\sim}{w}_0$ is m×1, $\underset{\sim}{H}$ is m×p_0, $\underset{\sim}{A}_0$ is $p_0 \times p_0$ and $\underset{\sim}{b}_0$ is $p_0 \times 1$. These quantities and $\underset{\sim}{y}'\underset{\sim}{y}$ are the only quantities needed to compute an iteration of The Iterative Procedure. Note that $\underset{\sim}{w}_0$, $\underset{\sim}{F}$, $\underset{\sim}{H}$, $\underset{\sim}{A}_0$, $\underset{\sim}{b}_0$ and $\underset{\sim}{y}'\underset{\sim}{y}$ may all be computed as the data are read in, thus eliminating any need to save the large matrix $\underset{\sim}{U}$. $\underset{\sim}{E}$ can then be formed at the start of each iteration and the matrices $\underset{\sim}{Q} = \underset{\sim}{E}^{-1}\underset{\sim}{F}$ and $\underset{\sim}{P} = \underset{\sim}{E}^{-1}\underset{\sim}{H}$ can be formed by solving linear equations. Thus it is never necessary to actually invert $\underset{\sim}{E}$.

The quantities necessary to perform one iteration of The Iterative Procedure are the elements of the matrix $\underset{\sim}{B}(\underset{\sim}{\sigma}_{(i)})$ and the vector $\underset{\sim}{c}[\underset{\sim}{\sigma}_{(i)}, \underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(i)})]$ defined in Section 5.4. The iteration required is then

$$\underset{\sim}{\sigma}_{(i+1)} = \underset{\sim}{B}^{-1}(\underset{\sim}{\sigma}_{(i)})\ c[\underset{\sim}{\sigma}_{(i)}, \underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(i)})],$$

which can also be solved without inverting the matrix $\underset{\sim}{B}(\underset{\sim}{\sigma}_{(i)})$. Thus it is never necessary to invert any matrix to use The Iterative Procedure. To obtain the elements of $\underset{\sim}{c}$, $\underset{\sim}{\alpha}$ must first be calculated from the equation

$$\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{X}\,\underset{\sim}{\alpha} = \underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{y}\ ,$$

or

$$\underset{\sim}{A}\,\underset{\sim}{\alpha} = \underset{\sim}{b}\ .$$

But

$$\underset{\sim}{A} = \underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{X}$$

$$= \frac{1}{\sigma_0}\underset{\sim}{X}'(\underset{\sim}{I} - \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')\underset{\sim}{X}$$

$$= \frac{1}{\sigma_0} (\underset{\sim}{X}'\underset{\sim}{X} - \underset{\sim}{X}'\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{X})$$

$$= \frac{1}{\sigma_0} (\underset{\sim}{A}_0 - \underset{\sim}{H}'\underset{\sim}{P}).$$

Also

$$\underset{\sim}{b} = \underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{y}$$

$$= \frac{1}{\sigma_0} \underset{\sim}{X}'(\underset{\sim}{I} - \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')\underset{\sim}{y}$$

$$= \frac{1}{\sigma_0} (\underset{\sim}{X}'\underset{\sim}{y} - \underset{\sim}{X}'\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{y})$$

$$= \frac{1}{\sigma_0} (\underset{\sim}{b}_0 - \underset{\sim}{P}'\ \underset{\sim}{w}_0)\ .$$

$\hat{\underset{\sim}{\alpha}}$ is obtained as $\hat{\underset{\sim}{\alpha}} = \underset{\sim}{A}^{-1}\underset{\sim}{b}$ and the elements of $\underset{\sim}{c}$ are calculated from the

$$[\underset{\sim}{c}]_i = (\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})'\ \underset{\sim}{\Sigma}^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}}),\quad i=0,1,\ldots,p_1.$$

For $i=1,2,\ldots,p_1$ this reduces to

$$(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})'\ \underset{\sim}{\Sigma}^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})$$

$$= \frac{1}{\sigma_0^2} (\underset{\sim}{y}-\underset{\sim}{X}\hat{\underset{\sim}{\alpha}})'(\underset{\sim}{I}-\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')\underset{\sim}{U}_i\underset{\sim}{U}_i'(\underset{\sim}{I}-\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')(\underset{\sim}{y}-\underset{\sim}{X}\hat{\underset{\sim}{\alpha}})$$

$$= \frac{1}{\sigma_0^2}\,[\underset{\sim}{U}_i'(\underset{\sim}{I}-\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')(\underset{\sim}{y}-\underset{\sim}{X}\hat{\underset{\sim}{\alpha}})]'[\underset{\sim}{U}_i'(\underset{\sim}{I}-\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')(\underset{\sim}{y}-\underset{\sim}{X}\hat{\underset{\sim}{\alpha}})]$$

$$\equiv \frac{1}{\sigma_0^2}\,\underset{\sim}{t}_i'\underset{\sim}{t}_i\ ,$$

where

$$\underset{\sim}{t}_i = \underset{\sim}{U}_i'(\underset{\sim}{I}-\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')(\underset{\sim}{y}-\underset{\sim}{X}\hat{\underset{\sim}{\alpha}})\ .$$

But if $\underset{\sim}{w}$ is defined by

$$\underset{\sim}{w} = \underset{\sim}{U}'\ (\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})$$

$$= \underset{\sim}{w}_0 - \underset{\sim}{H}\,\hat{\underset{\sim}{\alpha}}$$

$$= \begin{bmatrix} \underset{\sim}{w}_1 \\ \underset{\sim}{w}_2 \\ \vdots \\ \underset{\sim}{w}_{p_1} \end{bmatrix},$$

partitioned the same as $\underset{\sim}{w}_0$, then

$$\underset{\sim}{t}_i = \underset{\sim}{U}_i'(\underset{\sim}{I} - \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})$$

$$= \underset{\sim}{U}_i'(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}}) - \underset{\sim}{U}_i'\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})$$

$$= \underset{\sim}{w}_i - \sum_{j=1}^{p_1} \underset{\sim}{Q}_{ij}\underset{\sim}{w}_j,$$

where $\underset{\sim}{Q}$ and $\underset{\sim}{P}$ are partitioned just like $\underset{\sim}{E}$ and $\underset{\sim}{H}$ respectively. For i=0 it is necessary to calculate

$$(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})'\,\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})$$

$$= \frac{1}{\sigma_0^2}(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})'(\underset{\sim}{I} - \underset{\sim}{UE}^{-1}\underset{\sim}{U}')(\underset{\sim}{I} - \underset{\sim}{UE}^{-1}\underset{\sim}{U}')(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})$$

$$= \frac{1}{\sigma_0^2}\Big\{(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})'(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}}) - 2(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})'\underset{\sim}{UE}^{-1}\underset{\sim}{U}'(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})$$

$$+ (\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})'\,\underset{\sim}{UE}^{-1}\underset{\sim}{U}'\underset{\sim}{UE}^{-1}\underset{\sim}{U}'(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\underset{\sim}{\alpha}})\Big\}$$

$$= \frac{1}{\sigma_0^2}\Big\{\underset{\sim}{y}'\underset{\sim}{y} - 2\underset{\sim}{y}'\underset{\sim}{X}\hat{\underset{\sim}{\alpha}} + \hat{\underset{\sim}{\alpha}}'\underset{\sim}{X}'\underset{\sim}{X}\hat{\underset{\sim}{\alpha}} - 2[\underset{\sim}{U}'(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})]'\underset{\sim}{E}^{-1}\underset{\sim}{U}'(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})$$

$$+ [\underset{\sim}{E}^{-1}\underset{\sim}{U}'(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})]'\underset{\sim}{U}'\underset{\sim}{U}[\underset{\sim}{E}^{-1}\underset{\sim}{U}(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})]\Big\}$$

$$= \frac{1}{\sigma_0^2}\Big\{\underset{\sim}{y}'\underset{\sim}{y} - 2\underset{\sim}{b}_0'\,\hat{\underset{\sim}{\alpha}} + \hat{\underset{\sim}{\alpha}}\,\underset{\sim}{A}_0\,\hat{\underset{\sim}{\alpha}} - 2\underset{\sim}{w}'\underset{\sim}{z} + \underset{\sim}{w}'\underset{\sim}{F}\underset{\sim}{w}\Big\}$$

where $\underset{\sim}{z} \equiv \underset{\sim}{E}^{-1}\underset{\sim}{w}$ and $\underset{\sim}{z}$ is partitioned as $\underset{\sim}{w}$.

Now $[\underset{\sim}{B}]_{ij}$ must be computed, $i,j=0,1,\ldots,p_1$. For $i=j=0$,

$$[\underset{\sim}{B}]_{00} = \text{tr}\ \underset{\sim}{\Sigma}^{-1}\ \underset{\sim}{\Sigma}^{-1}$$

$$= \frac{1}{\sigma_0^2}\ \text{tr}(\underset{\sim}{I} - \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')(\underset{\sim}{I} - \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')$$

$$= \frac{1}{\sigma_0^2}\left\{\text{tr}\ \underset{\sim}{I} - 2\text{tr}\ \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}' + \text{tr}\ \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'\right\}$$

$$= \frac{1}{\sigma_0^2}\left\{n - 2\text{tr}\ \underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{U} + \text{tr}\ \underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{U}\right\}$$

$$= \frac{1}{\sigma_0^2}\left\{n - 2\text{tr}\ \underset{\sim}{Q} + \text{tr}\ \underset{\sim}{Q}^2\right\}.$$

For $j=0$, $i=1,2,\ldots,p_1$

$$[\underset{\sim}{B}]_{i0} = \text{tr}\ \underset{\sim}{\Sigma}^{-1}\ \underset{\sim}{\Sigma}^{-1}\ \underset{\sim}{G}_i$$

$$= \frac{1}{\sigma_0^2}\ \text{tr}(\underset{\sim}{I} - \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')(\underset{\sim}{I} - \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')\underset{\sim}{U}_i\underset{\sim}{U}_i'$$

$$= \frac{1}{\sigma_0^2}\left\{\text{tr}\ \underset{\sim}{U}_i\underset{\sim}{U}_i' - 2\text{tr}\ \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{U}_i\underset{\sim}{U}_i' + \text{tr}\ \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{U}_i\underset{\sim}{U}_i'\right\}$$

$$= \frac{1}{\sigma_0^2}\left\{n - 2\text{tr}\ \underset{\sim}{U}_i'\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{U}_i + \text{tr}\ \underset{\sim}{U}_i'\underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{U}_i\right\}$$

$$= \frac{1}{\sigma_0^2}\left\{n - 2\sum_{k=1}^{p_1} \operatorname{tr} \underset{\sim}{F}_{ik}\underset{\sim}{Q}_{ki} + \sum_{\ell=1}^{p_1}\sum_{k=1}^{p_1} \operatorname{tr} \underset{\sim}{Q}_{ik}\underset{\sim}{F}_{k\ell}\underset{\sim}{Q}_{\ell i}\right\} .$$

For $i,j=1,2,\ldots,p_1$

$$[\underset{\sim}{B}]_{ij} = \operatorname{tr} \underset{\sim}{\Sigma}^{-1}\underset{\sim}{G}_i\underset{\sim}{\Sigma}^{-1}\underset{\sim}{G}_j$$

$$= \frac{1}{\sigma_0^2} \operatorname{tr}(\underset{\sim}{I} - \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')\underset{\sim}{U}_i\underset{\sim}{U}_i'(\underset{\sim}{I} - \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')\underset{\sim}{U}_j\underset{\sim}{U}_j'$$

$$= \frac{1}{\sigma_0^2} \operatorname{tr} \underset{\sim}{U}_j'(\underset{\sim}{I} - \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')\underset{\sim}{U}_i\underset{\sim}{U}_i'(\underset{\sim}{I} - \underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}')\underset{\sim}{U}_j$$

$$= \frac{1}{\sigma_0^2} \operatorname{tr}[\underset{\sim}{U}_i'\underset{\sim}{U}_j - \underset{\sim}{U}_i'\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{U}_j]'[\underset{\sim}{U}_i'\underset{\sim}{U}_j - \underset{\sim}{U}_i'\underset{\sim}{U}\underset{\sim}{E}^{-1}\underset{\sim}{U}'\underset{\sim}{U}_j] ,$$

$$= \frac{1}{\sigma_0^2} \operatorname{tr}[\underset{\sim}{F}_{ij} - \sum_{k=1}^{p_1} \underset{\sim}{F}_{ik}\underset{\sim}{Q}_{kj}]'[\underset{\sim}{F}_{ij} - \sum_{k=1}^{p_1} \underset{\sim}{F}_{ik}\underset{\sim}{Q}_{kj}] .$$

The above formulae demonstrate that indeed only the matrices $\underset{\sim}{F}$, $\underset{\sim}{H}$, $\underset{\sim}{w}_0$, $\underset{\sim}{A}_0$ and $\underset{\sim}{b}_0$ described above need be carried in the calculations. Thus it is never necessary to save the large matrices $\underset{\sim}{U}$ and $\underset{\sim}{X}$. This enables problems of a reasonable size to be run using this program, which would not be possible if matrices of dimension n were needed.

The above algorithms are used in a sequence of subroutines which compute the requirements for each iteration. There is one more algebraic facet of the program of which the user should be aware. In Section 5.8 slowly converging sequences which oscillated above and below the final value were mentioned. This program has a feature designed to eliminate this problem. Whenever two iterates $\underset{\sim}{\sigma}_{(i-1)}$ and $\underset{\sim}{\sigma}_{(i)}$ are sufficiently different, $\underset{\sim}{\sigma}_{(i+1)}$ is formed by taking another iteration from $\underset{\sim}{\sigma}_{(i)}$ and then averaging. If averaging is necessary $\underset{\sim}{\sigma}^{(1)}$ is calculated from $\underset{\sim}{\sigma}_{(i)}$ as

$$\underset{\sim}{\sigma}^{(1)} = \underset{\sim}{B}^{-1}(\underset{\sim}{\sigma}_{(i)})\underset{\sim}{c}[\underset{\sim}{\sigma}_{(i)},\underset{\sim}{\alpha}(\underset{\sim}{\sigma}_{(i)})],$$

and then $\underset{\sim}{\sigma}_{(i+1)}$ is formed as

$$\underset{\sim}{\sigma}_{(i+1)} = \tfrac{1}{2}(\underset{\sim}{\sigma}_{(i)} + \underset{\sim}{\sigma}^{(1)}).$$

This feature eliminates the oscillating, but care must be taken that it not introduce a false convergence of its own. This is done by insisting that the last iteration must be one not involving the averaging process. The averaging and the safeguard process have worked very well on sample problems.

C.2. Setup and Output of the Computer Program

The computer program is designed to deliver as much freedom and convenience to the user as possible. The user must supply only two control cards. The first card contains the number of observations, n, the number of levels of fixed factors, p_0, the number of random factors, p_1, a number indicating how often the iterates are to be printed (see below) and two yes-no statements. The first is yes if user supplies initial guesses and no otherwise. The second is yes if the short cut notation for the $\underset{\sim}{U}$ matrix is used and no otherwise. (See below for explanation of the short cut notation.) The second control card contains the number of levels at which each random factor appears, m_i, $i=1,2,\ldots,p_1$. (Thus there are p_1 numbers on the second control card.)

After the control cards come the data cards. The data is read in by rows or observations. There are two cards per row. The first card contains the appropriate row of the $\underset{\sim}{U}$ matrix. This may be in the form of zeroes and ones (there will be $m \equiv \Sigma_{i=1}^{p_1} m_i$ numbers.) or in short cut notation. In short cut notation only p_1 numbers are used, each number stating in which column of $\underset{\sim}{U}_i$ the one appears; this is a unique description because by definition there is exactly one 1 in each row. The second card contains the appropriate row of the $\underset{\sim}{X}$ matrix and the observation on y. There must be n pairs of cards, one pair for each observation. The last card for a problem states yes or no--yes if another problem follows and no otherwise.

The output of the program consists first of the intermediate iterations the user asked to be printed. If the user placed a K on the first control card in the appropriate position, then the results after each K iterations are printed. Then follow the final results and the number of iterations required. The estimated large sample covariance matrices for $\hat{\underset{\sim}{\alpha}}$ and $\hat{\underset{\sim}{\sigma}}$ are also printed; these matrices are estimated by $[\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}(\hat{\underset{\sim}{\sigma}})\underset{\sim}{X}]^{-1}$ and $2\underset{\sim}{B}^{-1}(\hat{\underset{\sim}{\sigma}})$ respectively.

This computer program can handle fairly large problems with relative ease. The size of the largest matrix which should be inverted (or set of equations to be solved) is certainly less than 100 and probably closer to 50. (The reasons are twofold--computational accuracy and time requirements.) If $\underset{\sim}{\Sigma}$ were inverted directly the size of problem would be severely limited. However, using the indirect methods above much larger problems can be accommodated. Since very often n is approximately a multiple of the product of some of the m_i, even when the sum of the m_i is restricted to be small, n could be large. For instance, even if the sum of the m_i is less than 60, the possible values of n could be well over 1000. The simplicity of the control cards and data input makes the program easy to use even for the statistician who is unfamiliar with computer programming. It seems to be an effective and efficient program, judged by its performance on sample problems.

Copies of the program deck are available from the author upon request along with more detailed documentation and sample problems.

APPENDIX D

COMMENTS ON THE PROOF OF CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATES GIVEN BY HARTLEY AND RAO

In Section 1.1, it was mentioned that H. O. Hartley and J. N. K. Rao (1967) gave a proof of the consistency of the maximum likelihood estimates in the mixed model of the analysis of variance. It was remarked that the theorem was true but that the assumptions used were not all stated correctly, that the assumptions were restrictive and that important details were omitted from the proof. In this section some brief comments will be presented to elucidate these remarks. The notation used will be the notation used in this paper rather than that used by Hartley and Rao in order to preserve continuity.

The assumption stated incorrectly occurs on page 102 of the paper and is labeled 7.iii. It requires that for each design in the sequence of designs the matrix $\underset{\sim}{M} \equiv [\underset{\sim}{X}:\underset{\sim}{U}_1:\ldots:\underset{\sim}{U}_{p_1}]$ have a basis $\underset{\sim}{W} = [\underset{\sim}{X}:\underset{\sim}{U}^*]$ where $\underset{\sim}{U}^*$ contains at least one column from each $\underset{\sim}{U}_i$. An assumption of this sort is necessary to insure estimability of the parameters. That this assumption is not sufficient can be seen in the following counterexample. Let $\underset{\sim}{U}_1$ be any matrix such that the basis of $[\underset{\sim}{X}:\underset{\sim}{U}_1]$ is $[\underset{\sim}{X}:\underset{\sim}{U}^*]$ where $\underset{\sim}{U}^*$ contains at least p_1 columns from $\underset{\sim}{U}_1$. Then let $\underset{\sim}{U}_j = \underset{\sim}{U}_1$, $j=2,3,\ldots,p_1$. In such a model instead of each individual $\sigma_1,\sigma_2,\ldots,\sigma_{p_1}$ only $\sigma_1+\sigma_2+\ldots+\sigma_{p_1}$ can be estimated, yet this model satisfies the given assumption. One possible way to correctly state the assumption is to

use two separate assumptions like Assumptions 1.3.4 and 1.3.5.

The assumption which is restrictive is 7.i (p. 101) which requires that all the diagonal elements of the matrix $\underset{\sim}{U}_i' \underset{\sim}{U}_i$ (of which there are m_i) be less than or equal to some universal constant R for all n, $i=1,2,\ldots,p_1$. It was claimed in this paper that this required that the number of observations at any level of any random factor to be bounded and that it forced all normalizing sequences to be the order of n. That this assumption would be restrictive was shown in Section 1.1 by reference to Section 6.1. That the claims are true can be seen by the following argument. As Hartley and Rao point out, a diagonal element of $\underset{\sim}{U}_i' \underset{\sim}{U}_i$ represents exactly the number of observations at the appropriate level of the i^{th} factor and the sum of these elements must be n. (Assumption 1.3.6, which was also used by Hartley and Rao (p.95), can be restated to say that every observation is allocated to exactly one level of each factor and that each level is allocated at least one observation.) Thus there are m_i such elements, each less than R (which demonstrates the first claim) and greater than zero, and adding up to n; it follows that $m_i \geq n/R$, $i=1,2,\ldots,p_1$, which was to be shown.

In an attempt to show that the details of the proof of consistency omitted by Hartley and Rao may be important, a short outline of the method of proof will be attempted. The object is to apply the method of Wald (1949) and Wolfowitz (1949) to $[L(\underset{\sim}{y},\underset{\sim}{\theta})]^{\frac{1}{n}}$ where $L(\underset{\sim}{y},\underset{\sim}{\theta})$ is the likelihood function. One shows that for each $\eta > 0$ and $\epsilon > 0$ there is an $h \equiv h(\eta)$ and an $n_0 \equiv n_0(\eta,\epsilon)$ such that $h < 1$ and such that

$$P_0\left\{ \frac{\sup_{\underset{\sim}{\theta} \in \Omega(\eta)} L(\underset{\sim}{y},\underset{\sim}{\theta})}{L(\underset{\sim}{y},\underset{\sim}{\theta}_0)} > h^n \right\} < \epsilon ,$$

where $\Omega(\eta) = \{\underset{\sim}{\theta}: \|\underset{\sim}{\theta}-\underset{\sim}{\theta}_0\| > \eta\}$ and $\underset{\sim}{\theta}_0$ is the true parameter point. (That is, the likelihood function cannot be large unless $\underset{\sim}{\theta}$ is near $\underset{\sim}{\theta}_0$.) This is done by showing that

$$P_0\left\{ \sup_{\underset{\sim}{\theta} \in \Omega(\eta)} \frac{1}{n} \log L(\underset{\sim}{y},\underset{\sim}{\theta}) - \frac{1}{n} \log L(\underset{\sim}{y},\underset{\sim}{\theta}_0) > \log h \right\} < \epsilon .$$

The following are some of the things which must be shown in order to use the Wald-Wolfowitz argument.

(1) $\frac{1}{n} \log L(\underset{\sim}{y},\underset{\sim}{\theta}) \xrightarrow{p} \mathcal{E}_0[\frac{1}{n} \log L(\underset{\sim}{y},\underset{\sim}{\theta})]$ for all $\underset{\sim}{\theta}$.

(2) $\mathcal{E}_0[\frac{1}{n} \log L(\underset{\sim}{y},\underset{\sim}{\theta}) - \frac{1}{n} \log L(\underset{\sim}{y},\underset{\sim}{\theta}_0)] < 0$ for all n.

(3) $\lim_{n\to\infty} \mathcal{E}_0[\frac{1}{n} \log L(\underset{\sim}{y},\underset{\sim}{\theta}) - \frac{1}{n} \log L(\underset{\sim}{y},\underset{\sim}{\theta}_0)] < 0.$

(4) Continuity and limit conditions on $L(\underset{\sim}{y},\underset{\sim}{\theta})$.

Hartley and Rao prove (1) as Lemma 7.1 (p.102). They state that (2) and (4) follow from the assumptions; this is true, although a great deal of work is necessary to prove (4). They do not mention (3) at all. However, Condition (3) is very important. It is necessary that the expected values in (2) approach a strictly negative limit in order to insure that the probability of any particular difference of log likeli-

hoods being negative will converge to one. To see that (1) and (2) above are not enough, consider the sequence of random variables $Z_n \sim \mathcal{N}(-\frac{1}{n}, \frac{1}{n^2})$. Then $Z_n \xrightarrow{p} \mathcal{E}Z_n$ and $\mathcal{E}Z_n < 0$ for all n but $P\{Z_n < 0\} = 0.8413$ for all n. Thus a condition like Condition (3) must be proved; to prove this condition it is necessary to use arguments of the same type used in Section A.4.1 to show that the matrix $\underset{\sim}{J}$ was positive definite. It was shown in Section 1.1 that arguments about the positive definiteness of $\underset{\sim}{J}$ were closely related to problems of degenerating distributions. That such arguments must also be taken into account in this version of a consistency proof is entirely reasonable since this result is a stronger result than Theorem 4.4.1 and so this version should take into account any difficulties that arise there. Condition (3) can be proved using all the assumptions--including 7.i, but it is not true when different parameters converge at different rates. Thus this proof of consistency can be used under all of Hartley and Rao's assumptions but will not work in the more general cases because the limit in (3) degenerates to zero.

Interestingly enough, although the assumptions 7.i-7.iii are necessary to prove consistency, they are not necessary to prove (1). In fact the only necessities are that the elements of $\underset{\sim}{X}$ be bounded and that $\underset{\sim}{\theta}$ be an interior point of the parameter space. To prove (1) it suffices to prove, as Hartley and Rao did in Lemma 7.1 (p.102), that $\mathrm{Var}_0[\frac{1}{n} \log L(\underset{\sim}{y},\underset{\sim}{\theta})] = O(\frac{1}{n})$. But

$$\log L(\underset{\sim}{y},\underset{\sim}{\theta}) = \frac{n}{2}\log 2\pi - \frac{1}{2}\log|\underset{\sim}{\Sigma}| - \frac{1}{2}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\alpha})\ \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{y},\underset{\sim}{X}\underset{\sim}{\alpha}).$$

Therefore,

$$\mathrm{Var}_0[\log L(\underset{\sim}{y},\underset{\sim}{\theta})] = 2\mathrm{tr}(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\Sigma}_0)^2 + 4(\underset{\sim}{\alpha}-\underset{\sim}{\alpha}_0)\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\Sigma}_0\underset{\sim}{\Sigma}^{-1}\underset{\sim}{X}(\underset{\sim}{\alpha}-\underset{\sim}{\alpha}_0),$$

by Lemma B.1,

$$\leq 2n\lambda_{max}^2(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\Sigma}_0) + 4(\underset{\sim}{\alpha}-\underset{\sim}{\alpha}_0)'(\underset{\sim}{\alpha}-\underset{\sim}{\alpha}_0)\lambda_{max}(\underset{\sim}{X}\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X})$$

$$\cdot\lambda_{max}(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\Sigma}_0)^2$$

by Proposition A.3.4,

$$\leq 2\max_{i=0,1,\ldots,p_1}\left(\frac{\sigma_{0i}}{\sigma_i}\right)^2[n+2(\underset{\sim}{\alpha}-\underset{\sim}{\alpha}_0)'(\underset{\sim}{\alpha}-\underset{\sim}{\alpha}_0)$$

$$\lambda_{max}(\underset{\sim}{X}^{-1}\underset{\sim}{\Sigma}_0\underset{\sim}{X})]$$

by Lemma B.7. But

$$\lambda_{max}(\underset{\sim}{X}'\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{X}) \leq \lambda_{max}(\underset{\sim}{X}'\underset{\sim}{X})\ \lambda_{max}(\underset{\sim}{\Sigma}_0^{-1})$$

by Lemma B.8

$$= \lambda_{max}(\underset{\sim}{X}'\underset{\sim}{X})\ \frac{1}{\lambda_{min}(\underset{\sim}{\Sigma}_0)}$$

$$\leq \lambda_{max}(\underset{\sim}{X}'\underset{\sim}{X})\ \frac{1}{\sigma_{00}}$$

because $\lambda_{min}(\underset{\sim}{\Sigma}_0) \geq \sigma_{00}$. But if the elements of $\underset{\sim}{X}$ are bounded then $\lambda_{max}(\underset{\sim}{X}'\underset{\sim}{X}) \leq Kn$ for some constant K. Therefore

$$\mathrm{Var}_0[\frac{1}{n}\log L(\underset{\sim}{y},\underset{\sim}{\theta})] \le \frac{1}{n}[2 \max_{i=0,1,\ldots,p_1}\left(\frac{\sigma_{0i}}{\sigma_i}\right)^2 + 4(\underset{\sim}{\alpha}-\underset{\sim}{\alpha}_0)'(\underset{\sim}{\alpha}-\underset{\sim}{\alpha}_0)\cdot\frac{K}{\sigma_{00}}]$$

which was to be proved.

BIBLIOGRAPHY

Anderson, T. W. (1958), An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, Inc., New York.

Anderson, T. W. (1969), "Statistical inference for covariance matrices with linear structure", Proceedings of the Second International Symposium on Multivariate Analysis (P. R. Krishnaiah, ed.), Academic Press, Inc., New York, pp. 55-66.

Anderson, T. W. (1970), "Estimation of covariance matrices which are one linear combination or whose inverses are linear combinations of given matrices", Essays in Probability and Statistics, University of North Carolina Press, Chapel Hill, pp. 1-24.

Anderson, T. W. (1971a), The Statistical Analysis of Time Series, John Wiley & Sons, Inc., New York.

Anderson, T. W. (1971b), "Estimation of covariance matrices with linear structure and moving average processes of finite order", Technical Report No. 6, Stanford University.

Anderson, T. W. (1973), "Asymptotically efficient estimation of covariance matrices with linear structure", Annals of Statistics, **1**, pp. 135-141.

Anderson, T. W. and S. Das Gupta (1963), "Some inequalities on characteristic roots of matrices", Biometrika, **50**, pp. 522-523.

Buck, R. C. (1965), Advanced Calculus, McGraw-Hill Book Company, Inc., New York.

Cramér, H. (1946), Mathematical Methods of Statistics, Princeton University Press, Princeton.

Graybill, F. A. and R. A. Hultquist (1961), "Theories concerning Eisenharts Model II", Annals of Mathematical Statistics, 32, pp. 261-269.

Hartley, H. O. and J. N. K. Rao (1967), "Maximum likelihood estimation for the mixed analysis of variance model", Biometrika, 54, pp. 93-108.

Hartley, H. O. and S. R. Searle (1969), "On interaction variance components in mixed models", Biometrics, 25, pp. 573-6.

Hartley, H. O. and W. K. Vaughn (1972), "A computer program for the mixed analysis of variance model based on maximum likelihood", Statistical Papers in Honor of George W. Snedecor (T. A. Bancroft, Ed.),Iowa State University Press, Ames.

Harville, D. A. (1967a), "Estimability of variance components for the 2-way classification with interaction", Annals of Mathematical Statistics, 38, pp. 1508-19.

Henderson, C. R. (1953), "Estimation of variance and covariance components", Biometrics, 9, pp. 226-52.

Herbach, L. H. (1959), "Properties of Model II - type analysis of variance, A: optimum nature of the F-test for model II in the balanced case", Annals of Mathematical Statistics, 30, pp. 939-59.

Hultquist, R. A. and E. M. Atzinger (1972), "The mixed effects model and simultaneous diagonalization of symmetric matrices", Annals of Mathematical Statistics, 43, pp. 2024-30.

Hultquist, R. A. and F. A. Graybill (1965), "Minimal sufficient statistics for the 2-way classification mixed model design", Journal of the American Statistical Association, 60, pp. 182-92.

Isaacson, E. and H. B. Keller (1966), Analysis of Numerical Methods, John Wiley & Sons, Inc., New York.

Kale, B. K. (1961), "On the solution of the likelihood equations by iteration processes", Biometrika, 48, pp. 452-56.

Kale, B. K. (1962), "On the solution of likelihood equations by iteration processes, the multiparametric case", Biometrika, 49, pp. 479-86.

Koch, G. G. (1967a), "A general approach to the estimation of variance components", Technometrics, 9, pp. 93-118.

Koch, G. G. (1967b), "A procedure to estimate the population mean in random effect models", Technometrics, 9, pp. 577-86.

Koch, G. G. (1968), "Some further remarks on 'A general approach to the estimation of variance components'", Technometrics, 10, pp. 551-8.

Olkin, I. and S. J. Press (1969), "Testing and estimation for a stationary model", Annals of Mathematical Statistics, 40, pp. 1358-73.

Plackett, R. L. (1960), Principles of Regression Analysis, Oxford University Press, Oxford.

Ralston, A. (1965), A First Course in Numerical Analysis, McGraw-Hill Book Company, Inc., New York.

Rao, C. Radhakrishna (1965), Linear Statistical Inference and its Applications, John Wiley & Sons, Inc., New York.

Rao, C. Radhakrishna (1971), "Estimation of variance and covariance components - MINQUE theory", Journal of Multivariate Analysis, **1**, pp. 257-75.

Rao, C. Radhakrishna (1971), "Minimum variance quadratic unbiased estimation of variance components", Journal of Multivariate Analysis, **1**, pp. 445-56.

Rao, J. N. K. (1973), Personal Communication to T. W. Anderson.

Rudin, W. (1964), Principals of Mathematical Analysis (Second Edition), McGraw-Hill Book Company, Inc., New York.

Searle, S. R. (1970), "Large sample variances of maximum likelihood estimates of variance components", Biometrics, **26**, pp. 505-24.

Searle, S. R. (1971), "Topics in variance component estimation", Biometrics, **27**, pp. 1-76.

Scheffé, H. (1959), The Analysis of Variance, John Wiley & Sons, Inc., New York.

Silvey, S. D. (1961), "A note on maximum likelihood in the case of dependent random variables", Journal of the Royal Statistical Society (B), **23**, pp. 444-52.

Srivastava, J. N. (1966), "On testing hypotheses regarding a class of covariance structures", Psychometrika, **31**, pp. 147-64.

Srivastava, J. N. and R. L. Maik (1967), "On a new property of partially balanced association schemes useful in psychometric structural analysis", Psychometrika, **32**, pp. 279-89.

Thompson, W. A. Jr. (1962), "The problem of negative estimates of variance components", Annals of Mathematical Statistics, 33, pp. 273-89.

Vaughn, W. K. (1970), "A technique for maximum likelihood estimation in mixed models", Ph.D. thesis, Texas A&M University, College Station.

Wald, A. (1949), "Note on the consistency of the maximum likelihood estimates", Annals of Mathematical Statistics, 20, pp. 595-601.

Whitby, O. (1971), "Estimation of parameters in the generalized beta distribution", Technical Report No. 29, Stanford University.

Wilks, S. (1946), "Sample criteria for testing equality of means, equality of variances and equality of covariances in a normal multivariate distribution", Annals of Mathematical Statistics, 17, pp. 257-281.

Wolfowitz, J. (1949), "On Wald's proof of the consistency of the maximum likelihood estimate", Annals of Mathematical Statistics, 20, pp. 601-2.

Woodbury, M. (1950), "Inverting Modified Matrices", Memorandum Report No. 42, Statistical Research Group, Princeton.